Exploring the Limits of Graph Invariant- and Spectrum-Based Discrimination of (Sub)structures‡

Christoph Rücker,* Gerta Rücker, and Markus Meringer

Department of Mathematics, Universität Bayreuth, D-95440 Bayreuth, Germany

Received December 1, 2001

The limits of a recently proposed computer method for finding all distinct substructures of a chemical structure are systematically explored within comprehensive graph samples which serve as supersets of the graphs corresponding to saturated hydrocarbons, both acyclic (up to $n = 20$) and (poly)cyclic (up to $n = 10$). Several pairs of smallest graphs and compounds are identified that cannot be distinguished using selected combinations of invariants such as combinations of Balaban's index J and graph matrix eigenvalues. As the most important result, it can now be stated that the computer program NIMSG, using J and distance eigenvalues, is safe within the domain of mono- through tetracyclic saturated hydrocarbon substructures up to $n = 10$ (oligocyclic decanes) and of all acyclic alkane substructures up to $n = 19$ (nonadecanes), i.e., it will not miss any of these substructures. For the regions surrounding this safe domain, upper limits are found for the numbers of substructures that may be lost in the worst case, and these are low. This taken together means that the computer program can be reasonably employed in chemistry whenever one is interested in finding the saturated hydrocarbon substructures. As to unsaturated and heteroatom containing substructures, there are reasons to conjecture that the method's resolving power for them is similar.

INTRODUCTION

Substructures and subgraphs of chemical structures are becoming increasingly important in description of chemical compounds' properties and reactivity,[1a] in similarity and complexity considerations,[1b,c] in physical and biological property prediction,[1d] and in automatic structure elucidation from spectral data.[2] We recently developed computer programs capable of finding all connected subgraphs in simple graphs,[3] all connected substructures and distinct connected substructures in colored multigraphs and chemical structures,[4] and all connected substructures and subgraphs and distinct connected substructures and subgraphs in colored multigraphs and chemical structures.[5] In such an endeavor the ability to distinguish very similar graphs is obviously a central issue and is in fact the limiting factor. Since a fast computer method for reliably discriminating all nonisomorphic graphs was not at our hands, the best we could do was to use graph invariants of discriminating power as high as possible.

A graph herein is understood to be simple, connected, and undirected. It contains n vertices, m edges, and $c = m - n + 1$ cycles. A graph invariant is a number calculated for a graph from its structure according to a well-defined procedure, its value is independent of how the graph is drawn or how its vertices are numbered. Being a simple number, a graph invariant carries less information than the graph itself, and this loss of information results in graph invariants being more or less degenerate, i.e., nonisomorphic graphs may have the same value of a particular invariant.

An easy-to-calculate graph invariant which is nevertheless considered rather well-discriminating is Balaban's index J.[6] Index J is of low degeneracy (has high discriminating power) compared to several other well-known graph invariants, in that the smallest J-equivalent simple graphs have $n = 6$ vertices, the smallest J-equivalent tree (=acyclic) graphs are found in the $n = 10$ family, and the smallest J-equivalent alkanes (4-trees) are dodecanes.[7]

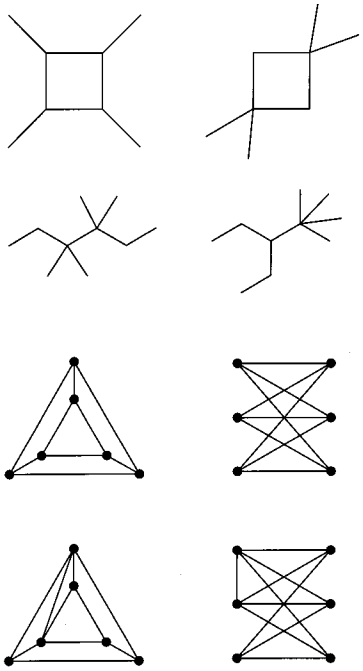

Figure 1. Pairs of J-equivalent but not isospectral graphs.

A better resolution should be achievable by using, instead of one graph invariant, a sequence of several graph invariants,[8] such as a spectrum. A graph's spectrum is the sequence of the eigenvalues of its adjacency matrix, a one-dimensional array of n graph invariants. The spectrum still contains less information than the graph itself, i.e., two nonisomorphic graphs may exhibit the same spectrum, in which case they

* Corresponding author phone: +49-921-553386; fax: +49-921-553385; e-mail: Christoph.Ruecker@uni-bayreuth.de.

‡ Dedicated to Professor A. T. Balaban on the occasion of his 71st birthday.

10.1021/ci010121y CCC: $22.00
Published on Web 04/12/2002

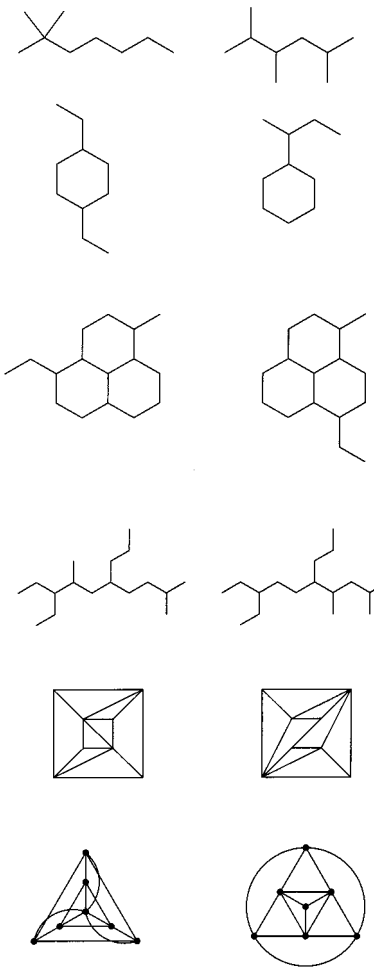

Figure 2. Pairs of isospectral but not *J*-equivalent graphs. The last four pairs are even distance-isospectral.

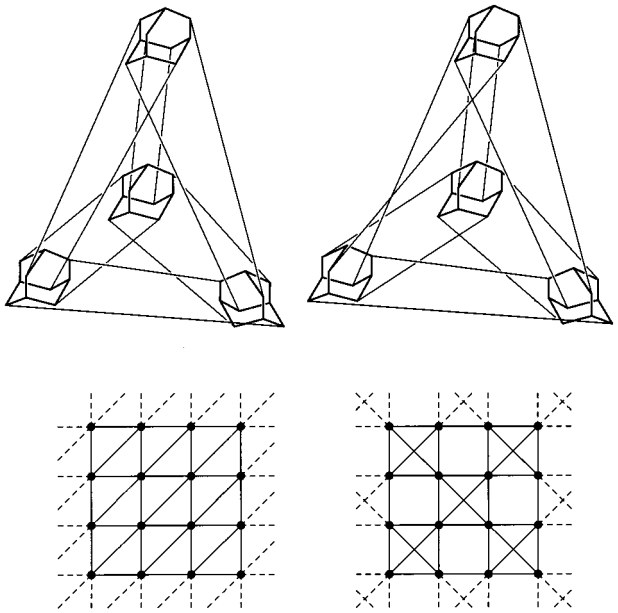

Figure 3. Pairs of graphs that are both *J*-equivalent and isospectral (and distance-isospectral).

are called isospectral or cospectral graphs. The smallest isospectral connected simple graphs have $n = 6$ vertices,[9,10] the smallest isospectral tree graphs are in the $n = 8$ family,[9] and the smallest isospectral alkanes are nonanes.[11] These numerical results, when compared to those for index *J* cited above, are somewhat unexpected, they emphasize the extraordinarily high resolving power of *J*. Whether or not index *J* is generally better resolving than the spectrum was never investigated. It was, however, proven that for increasing *n* the fraction of isospectral trees among all trees approaches 1, i.e., "almost all trees are cospectral".[12]

Table 1. Numbers of Graphs with *n* Vertices and *m* Edges for $n \leq 10$

m \ n	1	2	3	4	5	6	7	8	9	10
0	1									
1		1								
2			1							
3			1	2						
4				2	3					
5				1	5	6				
6				1	5	13	11			
7					4	19	33	23		
8					2	22	67	89	47	
9					1	20	107	236	240	106
10					1	14	132	486	797	657
11						9	138	814	2075	2678
12						5	126	1169	4495	8548
13						2	95	1454	8404	22950
14						1	64	1579	13855	53863
15						1	40	1515	20303	112618
16							21	1290	26631	211866
17							10	970	31400	361342
18							5	658	33366	561106
19							2	400	31996	795630
20							1	220	27764	1032754
21							1	114	21817	1229228
22								56	15558	1343120
23								24	10096	1348674
24								11	5984	1245369
25								5	3247	1057896
26								2	1635	827086
27								1	770	595418
28								1	344	394820
29									148	241428
30									63	136370
31									25	71293
32									11	34652
33									5	15767
34									2	6757
35									1	2768
36									1	1102
37										428
38										165
39										66
40										26
41										11
42										5
43										2
44										1
45										1
Σ	1	1	2	6	21	112	853	11117	261080	11716571

More discriminant than the usual (adjacency matrix) spectrum seems to be the graph distance spectrum, i.e., the sequence of eigenvalues of the graph distance matrix:[13] The smallest distance-isospectral trees have $n = 17$ vertices and are alkane (heptadecane) graphs,[14,15] while the smallest distance-isospectral simple graphs were not known at the beginning of this study. So neither simple-number graph invariants nor spectra seem to uniquely characterize a graph, i.e., discriminate it from all nonisomorphic graphs.

We had found that as a rule of thumb pairs of *J*-equivalent graphs are discriminated by their adjacency or distance spectra (see Figure 1), and conversely typical isospectral and even distance-isospectral graphs are discriminated by their *J* values (see Figure 2). So we formulated the working hypothesis that this will be generally the case, at least for small and not too complex (molecular) graphs. Accordingly, we decided to use for graph discrimination in our computer program NIMSG the combination of *J* and adjacency or distance eigenvalues.[4]

Of course a pair of graphs that are at the same time *J*-equivalent *and* isospectral cannot be distinguished by this method. Thus if two such graphs appear both as subgraphs

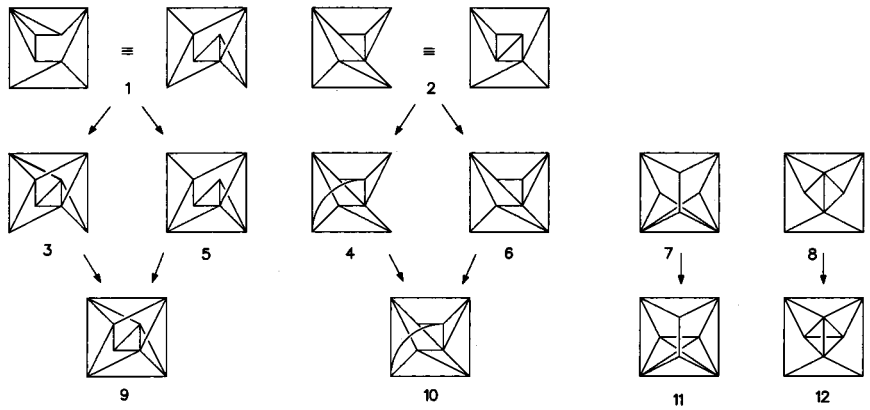

Figure 4. Smallest graphs that are pairwise both *J*-equivalent and isospectral (and moreover distance-isospectral).

Table 2. Numbers of Distinct *J* Values and *Resolution* by *J* for Graphs of $n \leq 10$

n	1	2	3	4	5	6		7		8		9		10	
m															
0	1														
1		1													
2			1												
3			1	2											
4				2	3										
5				1	5	6	1								
6				1	5	13	1	11	1						
7					4	19	1	33	1	23	1				
8					2	21	0.955	67	1	88	0.989	47	1		
9					1	17	**0.850**	107	1	235	0.996	240	1	105	0.991
10					1	13	0.929	128	0.970	480	0.988	792	0.994	644	0.980
11						9	1	121	0.877	793	0.974	2058	0.992	2626	0.980
12						5	1	100	0.794	1088	0.931	4417	0.983	8377	0.980
13						2	1	72	**0.758**	1239	0.852	8085	0.962	22421	0.977
14						1	1	50	0.781	1160	0.735	12694	0.916	51997	0.965
15						1	1	34	0.850	956	0.631	16875	0.831	105918	0.941
16								20	0.952	727	0.564	18923	0.711	189264	0.893
17								10	1	513	**0.529**	18419	0.587	294705	0.816
18								5	1	351	0.533	15940	0.478	394460	0.703
19								2	1	229	0.573	12560	0.393	454946	0.571
20								1	1	142	0.645	9238	0.333	459506	0.445
21								1	1	85	0.746	6493	0.298	416763	0.339
22										48	0.857	4439	**0.285**	348105	0.259
23										23	0.958	2937	0.291	272470	0.202
24										11	1	1883	0.315	202590	0.163
25										5	1	1172	0.361	144260	0.136
26										2	1	705	0.431	99228	0.120
27										1	1	401	0.521	66345	0.111
28										1	1	219	0.637	43425	**0.110**
29												113	0.764	27863	0.115
30												55	0.873	17586	0.129
31												24	0.960	10833	0.152
32												11	1	6505	0.188
33												5	1	3769	0.239
34												2	1	2089	0.309
35												1	1	1111	0.401
36												1	1	569	0.516
37														276	0.648
38														128	0.776
39														58	0.879
40														25	0.962
41														11	1
42														5	1
43														2	1
44														1	1
45														1	1
	1	1	2	6	21	107	0.955	762	0.893	8200	0.738	138749	0.531	3648987	0.311

in a graph, the result will be a wrong (low by 1) number of distinct subgraphs. Before the present study was initiated we knew of only a few such pairs of graphs, e.g. two regular cubic (degree of each vertex equals 3) simple graphs of 40 vertices[16] or two nonmolecular graphs of 16 vertices (all vertex degrees equal to 6).[17] These graphs are shown in Figure 3, and further examples can be found in the work of Weisfeiler[18] and Mathon.[19] In the context of molecular structures all these graphs seemed irrelevant, most for their high vertex degrees (>4), the first-mentioned pair for their size in combination with their regularity and the unfavorable geometry of any 3D-realization.

Treating the complete graphs K_n up to $n = 7$ as tests of our program NIMSG had resulted in the correct numbers of distinct subgraphs.[20] So we knew that at least up to and including $n = 7$ no such "dangerous" pairs of simultaneously *J*-equivalent and isospectral simple graphs exist. The aim of the present work was to find out whether such dangerous pairs are a realistic threat in finding molecular substructures, in particular in the application of NIMSG to molecular struc-

Table 3. Numbers of Distinct Adjacency Spectra and *Resolution* by Adjacency Spectra for Graphs of $n \leq 10$

n	1	2	3	4	5	6		7		8		9		10	
m															
0	1														
1		1													
2			1												
3			1	2											
4				2	3										
5				1	5	6	1								
6				1	5	13	1	11	1						
7					4	18	**0.947**	33	1	22	0.957				
8					2	22	1	63	**0.940**	84	0.944	42	0.894		
9					1	20	1	103	0.963	225	0.953	224	0.933	102	0.962
10					1	14	1	125	0.947	445	**0.916**	706	0.886	596	0.907
11						9	1	132	0.957	764	0.939	1895	0.913	2447	0.914
12						5	1	120	0.952	1078	0.922	4013	0.893	7562	0.885
13						2	1	93	0.979	1352	0.930	7613	0.906	20590	0.897
14						1	1	63	0.984	1470	0.931	12403	**0.895**	47500	0.882
15						1	1	39	0.975	1412	0.932	18280	0.900	100231	0.890
16								20	0.952	1213	0.940	23844	0.895	186998	0.883
17								10	1	923	0.952	28175	0.897	320119	0.886
18								5	1	625	0.950	29995	0.899	494753	0.882
19								2	1	385	0.963	28883	0.903	702885	0.883
20								1	1	214	0.973	25106	0.904	910360	**0.881**
21								1	1	112	0.982	19874	0.911	1084029	0.882
22										55	0.982	14234	0.915	1185376	0.883
23										24	1	9307	0.922	1191702	0.884
24										11	1	5554	0.928	1101401	0.884
25										5	1	3045	0.938	937402	0.886
26										2	1	1548	0.947	735474	0.889
27										1	1	738	0.958	530739	0.891
28										1	1	334	0.971	353419	0.895
29												144	0.973	216990	0.899
30												62	0.984	123286	0.904
31												25	1	64811	0.909
32												11	1	31765	0.917
33												5	1	14586	0.925
34												2	1	6320	0.935
35												1	1	2614	0.944
36												1	1	1053	0.956
37														414	0.967
38														162	0.982
39														65	0.985
40														26	1
41														11	1
42														5	1
43														2	1
44														1	1
45														1	1
	1	1	2	6	21	111	0.991	821	0.962	10423	0.938	236064	0.904	10375797	0.88

tures. This was to be done by systematically identifying the smallest such pairs of graphs.

RESULTS AND DISCUSSION

So questions arose as to the size and identities of the smallest simple graphs simultaneously being *J*-equivalent *and* isospectral and to the nature of such graphs — "molecular" or not. Unfortunately, no simple definition of a molecular graph is available. Therefore in the following we treat the sets of connected simple graphs, of connected simple 4-graphs, of trees and of 4-trees up to a certain vertex number, each of which is a superset of cyclic or acyclic saturated hydrocarbon graphs, respectively.

New hardware now allowed us to fully treat the complete graph K_8. As it happened, the number of distinct connected subgraphs of $n = 8$ found was low, 11111 instead of 11117,[20] even if all eight adjacency matrix eigenvalues or all eight distance matrix eigenvalues were used together with *J* for discrimination (instead of the routinely employed two adjacency or two distance eigenvalues). In detail, our procedure found 1578 instead of 1579 distinct connected simple graphs of $n = 8$, $m = 14$ (corresponding to heptacyclic octanes), 1512 instead of 1515 distinct connected simple graphs of $n = 8$, $m = 15$ (octacyclic octanes), and 1288 instead of 1290 distinct connected simple graphs of $n = 8$, $m = 16$ (nonacyclic octanes). Each distinct subgraph found occurs in many copies within K_8 due to its high symmetry, e.g. a typical occurrence number of $n = 8$, $m = 16$ subgraphs in K_8 is 23040. For $n = 8$, all other m ($7 \leq m \leq 13$ and $17 \leq m \leq 28$), the numbers of distinct connected simple graphs found were correct.[20] At this stage we knew that there must exist a few pairs of graphs with the sought-after combination of properties for $n = 8$, $m = 14-16$, but so far it was impossible to identify them. Comfortably, it was also clear that hepta-, octa-, and nonacyclic graphs of eight vertices are not molecular graphs.[21]

The key to the successful identification reported here is a complete generation free of redundancy of all connected simple graphs of $n = 8$, $m = 14$, 15, 16, that was now performed using MOLGEN 4.0.[22] Within MOLGEN 4.0, isomorphic graphs are identified, and nonisomorphic graphs are distinguished by a canonical numbering scheme. Calculation of *J* and the eigenvalues for all 1579 graphs of $n = 8$, $m = 14$, 1515 graphs of $n = 8$, $m = 15$, and 1290 graphs of $n = 8$, $m = 16$ and sorting by *J* or/and the eigenvalues within each class led to the following observations:[23]

(i) There are many pairs, triplets, and higher tuples of *J*-equivalent graphs in each of these classes.[24]

(ii) There are many pairs and several triplets of isospectral and even distance isospectral graphs in these graph classes.

Table 4. Numbers of Distinct Distance Spectra and *Resolution* by Distance Spectra for Graphs of $n \leq 10$

m \ n	1	2	3	4	5	6	7		8		9		10	
0	1													
1		1												
2			1											
3			1	2										
4				2	3									
5				1	5	6								
6				1	5	13	11	1						
7					4	19	33	1	23	1				
8					2	22	67	1	89	1	47	1		
9					1	20	107	1	236	1	240	1	106	1
10					1	14	131	0.992	484	0.996	796	0.999	657	1
11						9	135	0.978	810	0.995	2069	0.997	2676	0.999
12						5	123	0.976	1154	0.987	4473	0.995	8528	0.998
13						2	93	0.979	1423	0.979	8333	0.992	22880	0.997
14						1	63	0.984	1523	0.965	13679	0.987	53586	0.995
15						1	39	0.975	1446	0.954	19911	0.981	111796	0.993
16							21	1	1227	**0.951**	25852	0.971	209671	0.990
17							10	1	928	0.957	30128	0.959	356153	0.986
18							5	1	632	0.960	31556	0.946	550142	0.980
19							2	1	385	0.963	29855	0.933	774502	0.973
20							1	1	214	0.973	25690	0.925	995890	0.964
21							1	1	112	0.982	20137	**0.923**	1171682	0.953
22									55	0.982	14367	0.923	1263321	0.937
23									24	1	9352	0.926	1251178	0.928
24									11	1	5572	0.931	1140892	0.916
25									5	1	3052	0.940	960432	0.908
26									2	1	1550	0.948	746945	0.903
27									1	1	739	0.960	535900	**0.900**
28									1	1	334	0.971	355429	0.900
29											145	0.980	217714	0.902
30											62	0.984	123533	0.906
31											25	1	64909	0.910
32											11	1	31798	0.918
33											5	1	14591	0.925
34											2	1	6325	0.936
35											1	1	2619	0.946
36											1	1	1057	0.959
37													416	0.972
38													162	0.982
39													65	0.985
40													26	1
41													11	1
42													5	1
43													2	1
44													1	1
45													1	1
	1	1	2	6	21	112	842	0.987	10785	0.970	247984	0.950	10975530	0.937

(iii) For $n = 8$, there are exactly the following six pairs of graphs which are simultaneously *J*-equivalent and isospectral:

$m = 14$ class (heptacyclic octanes): **1** and **2** shown in Figure 4;

$m = 15$ class (octacyclic octanes): **3** and **4**, **5** and **6**, and **7** and **8**;

$m = 16$ class (nonacyclic octanes): **9** and **10** and **11** and **12**.

Four of these are planar graphs (**1,2,6,8**), the others are nonplanar.

Surprisingly, within each such pair of graphs even the distance matrix eigenvalues coincide, i.e., these graphs are pairwise not only *J*-equivalent and isospectral, but even distance-isospectral.[25] Furthermore, the Wiener index W and Hosoya index *Z* values (and their building blocks p(G,k)) pairwise coincide. In fact these graphs are "topological twin graphs" in the sense of Hosoya,[26] but being *J*-equivalent they are even more similar to one another than required by the definition of topological twins.[27] Furthermore, with respect to the number of edges graphs **1** and **2** are smaller than Hosoya's smallest topological twins.

These graphs are genetically related, they form two families: From **1** both **3** and **5** can be formed by addition of an edge and adding the respective other edge to either of these results in **9**. Likewise, from **2** (the twin of **1**) by adding an edge **4** and **6** (the twins of **3** and **5**) can be formed, and either of these leads to **10** (the twin of **9**) by adding the respective other edge. The second family is formed by **7**, its twin **8**, and **11** and its twin **12**, where the latter result from addition of an edge to either of the former. As anticipated, all these graphs are nonmolecular graphs due to vertex degrees exceeding 4.

Thus a partial answer to the question on the limits of validity of our working hypothesis above was found. However, from these findings the following questions arose: Are distance spectra really more discriminating than adjacency spectra? How frequent are *J*-equivalent graphs, isospectral graphs, distance-isospectral graphs, graphs both *J*-equivalent and isospectral, graphs both *J*-equivalent and distance-isospectral among the graphs of $n > 8$, and in particular among the molecular graphs of that size? What are the smallest molecular graphs simultaneously *J*-equivalent and isospectral/distance-isospectral?

Graphs of $n \leq 10$. To get an idea on possible answers to these questions we decided first to systematically look for degeneracies in *J* and adjacency and distance spectra within the set of all connected simple graphs of up to $n = 10$, which using MOLGEN 4.0 seemed to be a realistic task.

Thus all connected simple graphs of 1, 2, . . . , 10 vertices (nearly 12 million graphs) were generated using MOLGEN 4.0 in classes of constant numbers of vertices and edges, their *J* values and adjacency and distance spectra were calculated, the numbers of distinct *J* values and distinct

Table 5. Numbers of Graphs with Distinct Combination of *J* and Adjacency Spectrum and *Resolution* by This Combination for $n \leq 10$

n	1	2	3	4	5	6	7	8		9		10	
m													
0	1												
1		1											
2			1										
3			1	2									
4				2	3								
5				1	5	6							
6				1	5	13	11						
7					4	19	33	23	1				
8					2	22	67	89	1	47	1		
9					1	20	107	236	1	240	1	106	1
10					1	14	132	486	1	797	1	657	1
11						9	138	814	1	2075	1	2678	1
12						5	126	1169	1	4494	0.999	8548	1
13						2	95	1454	1	8401	0.999	22949	0.999
14						1	64	1578	0.999	13849	0.999	53854	0.999
15						1	40	1512	**0.998**	20282	0.999	112587	0.999
16							21	1288	0.998	26566	0.998	211763	0.999
17							10	970	1	31268	0.996	361019	0.999
18							5	658	1	33163	0.994	560215	0.998
19							2	400	1	31727	0.992	793471	0.997
20							1	220	1	27505	**0.991**	1028231	0.996
21							1	114	1	21647	0.992	1221366	0.994
22								56	1	15442	0.993	1331659	0.991
23								24	1	10036	0.994	1334090	0.989
24								11	1	5957	0.995	1229374	0.987
25								5	1	3238	0.997	1043152	0.986
26								2	1	1633	0.999	815323	0.986
27								1	1	770	1	586926	**0.986**
28								1	1	344	1	389474	0.986
29										148	1	238458	0.988
30										63	1	134959	0.990
31										25	1	70681	0.991
32										11	1	34437	0.994
33										5	1	15697	0.996
34										2	1	6741	0.998
35										1	1	2765	0.999
36										1	1	1102	1
37												428	1
38												165	1
39												66	1
40												26	1
41												11	1
42												5	1
43												2	1
44												1	1
45												1	1
	1	1	2	6	21	112	853	11111	0.999	259737	0.995	11612987	0.991

adjacency and distance spectra were determined via sorting by *J* or the eigenvalues, respectively (two spectra are distinct if they differ in at least one eigenvalue). The results are shown in Tables 1−6. In the tables every fifth row is underlined for better orientation. Table 1 gives the numbers of connected simple graphs in classes of constant *n* and *m*, as known[20] and as generated by MOLGEN 4.0. These numbers serve as reference values against which to compare the entries in Tables 2−6. Table 2 gives the numbers of distinct *J* values within each *n*,*m*-class,[28] Tables 3 and 4 show the numbers of distinct adjacency spectra and distinct distance spectra. Tables 5 and 6 give the numbers of distinct combinations of *J* and adjacency spectra and of *J* and distance spectra, respectively, for the same classes of graphs. Tables 2−6 also show *in italics* the resolution of the respective graph invariant (combination), i.e., each italic entry is the entry left to it divided by the corresponding entry in Table 1. In the tables the "dangerous" region, the range where the particular invariant (combination) cannot uniquely characterize all graphs, is shaded.

Tables 2−6 all give qualitatively the same picture: The resolutions (discriminating powers) of the graph invariants gradually drop for increasing *n*. For increasing *m* within each *n* the discriminating powers initially drop and then pass through a minimum (printed in bold), finally approaching 1 again. The latter feature is explained by the fact that for increasing *m* the numbers of distinct graphs first increase but then decrease again until the second-highest and highest *m* classes contain only one graph each, the K_n-minus-an-edge and K_n graphs, so degeneracy in these classes cannot exist.

Huge differences are seen in the discriminating power of the graph invariants considered here:

(i) Index *J* is very good for acyclic and oligocyclic graphs (the few first entries in each column), i.e., the domain of real molecular species. In that region *J* is even better than the adjacency spectrum. However, down the columns, i.e., for polycyclic graphs, *J*'s resolution sharply drops, so that most graphs are better resolved by their spectra.

The different behavior of *J* for acyclic and polycyclic graphs may be understood: *J* exploits the differences in the (topological) distances between vertices in a graph, more exactly the differences between the distance sums. In going from a tree to a polycycle, long distances are replaced by shorter ones, those that are present in any graph. In the extreme case, the K_n, all distances are 1 and all distance sums equal $n-1$. So in that direction the distances (and their sums) tend to equalize for the vertices in a graph and between isomeric graphs as well.

(ii) The distance spectrum is always at least as discriminating as the adjacency spectrum.

(iii) The combinations of *J* and spectra, particularly *J* and the distance spectrum, are unrivalled, as expected.[29] In Table 6 in each column the first five resolution entries are 1, that is, the domain of acyclic to tetracyclic graphs (saturated

Table 6. Numbers of Graphs with Distinct Combination of *J* and Distance Spectrum and *Resolution* by This Combination for $n \leq 10$

n	1	2	3	4	5	6	7	8		9		10	
m													
0	1												
1		1											
2			1										
3			1	2									
4				2	3								
5				1	5	6							
6				1	5	13	11						
7					4	19	33	23	1				
8					2	22	67	89	1	47	1		
9					1	20	107	236	1	240	1	106	1
10					1	14	132	486	1	797	1	657	1
11						9	138	814	1	2075	1	2678	1
12						5	126	1169	1	4495	1	8548	1
13						2	95	1454	1	8403	0.999	22950	1
14						1	64	1578	0.999	13853	0.999	53862	0.999
15						1	40	1512	**0.998**	20296	0.999	112612	0.999
16							21	1288	0.998	26596	0.999	211837	0.999
17							10	970	1	31313	0.997	361226	0.999
18							5	658	1	33205	0.995	560691	0.999
19							2	400	1	31753	0.992	794373	0.998
20							1	220	1	27513	**0.991**	1029620	0.997
21							1	114	1	21648	0.992	1223130	0.995
22								56	1	15442	0.993	1333361	0.993
23								24	1	10036	0.994	1335375	0.990
24								11	1	5957	0.995	1230147	0.988
25								5	1	3238	0.997	1043527	0.986
26								2	1	1633	0.999	815449	0.986
27								1	1	770	1	586960	**0.986**
28								1	1	344	1	389484	0.986
29										148	1	238460	0.988
30										63	1	134959	0.990
31										25	1	70681	0.991
32										11	1	34437	0.994
33										5	1	15697	0.996
34										2	1	6741	0.998
35										1	1	2765	0.999
36										1	1	1102	1
37												428	1
38												165	1
39												66	1
40												26	1
41												11	1
42												5	1
43												2	1
44												1	1
45												1	1
	1	1	2	6	21	112	853	11111	0.999	259910	0.996	11622140	0.992

hydrocarbons) of up to $n = 10$ is "safe" if the combination of *J* and distance spectrum is used for discrimination.

4-Graphs of $n \leq 10$. Program NIMSG for finding distinct substructures was developed primarily for chemistry, where one is mostly interested in acyclic through oligocyclic graphs (e.g. for $n \leq 10$, $n - 1 \leq m \leq \sim n + 5$) and in particular in graphs of vertex degrees not exceeding 4, the valency of carbon (so-called 4-graphs). The above procedure was therefore repeated for simple connected 4-graphs, the graph sample most closely approximating the acyclic and oligo- through polycyclic saturated hydrocarbons, up to $n = 10$. The results are shown in Tables 7−12. Contrary to naïve expectation, the resolution in this sample is not decisively better than in the sample of all graphs, so that the resolution problem essentially remains the same. Though the numbers of 4-graphs (Table 7) are lower than those of all graphs (Table 1), and often far lower, particularly in higher *m* classes, the 4-graphs are a more uniform group, so finding differences among them is more difficult. Index *J* suffers most from this fact.

The observations made in the sample of all graphs are reproduced in the 4-graphs. From Table 12 it again (and necessarily) follows that all acyclic to tetracyclic saturated hydrocarbons of up to at least $n = 10$ are distinguished by the combination of *J* and distance spectrum.

The pair of smallest *J*-equivalent *and* isospectral 4-graphs was identified in the $n = 9$, $m = 12$ class (tetracyclic nonanes, Table 11), graphs **13** and **14**; the pair of smallest

Table 7. Numbers of 4-Graphs with *n* Vertices and *m* Edges for $n \leq 10$

n	1	2	3	4	5	6	7	8	9	10
m										
0	1									
1		1								
2			1							
3			1	2						
4				2	3					
5				1	5	5				
6				1	5	12	9			
7					4	17	29	18		
8					2	18	56	73	35	
9					1	14	79	182	185	75
10					1	8	79	326	573	475
11						3	59	430	1278	1792
12						1	31	427	2161	4875
13							9	298	2768	10162
14							2	134	2616	16461
15								35	1714	20346
16								6	707	18436
17									154	11477
18									16	4399
19										845
20										59
Σ	1	1	2	6	21	78	353	1929	12207	89402

J-equivalent *and* distance-isospectral 4-graphs was found in the $n = 9$, $m = 13$ class (pentacyclic nonanes, Table 12), graphs **15** and **16**. These graphs are shown in Figure 5. Though as 4-graphs they fulfill the formal condition for molecular graphs and though they are planar graphs, a chemist will doubt the viability of their molecular counterparts, due to their presumably extremely strained nature: No reasonable geometric structures (having usual bond lengths,

Table 8. Numbers of Distinct *J* Values and *Resolution* by *J* for 4-Graphs of $n \leq 10$

n	1	2	3	4	5	6		7		8		9		10	
m															
0	1														
1		1													
2			1												
3			1	2											
4				2	3										
5				1	5	5	1								
6				1	5	12	1	9	1						
7					4	17	1	29	1	18	1				
8					2	17	0.944	56	1	72	0.986	35	1		
9					1	11	**0.786**	79	1	181	0.995	185	1	75	1
10					1	7	0.875	75	0.949	320	0.982	570	0.995	468	0.986
11						3	1	47	0.797	414	0.963	1271	0.995	1767	0.986
12						1	1	18	0.581	379	0.888	2122	0.982	4799	0.984
13								3	**0.333**	214	0.718	2612	0.944	9966	0.981
14								1	0.500	57	0.425	2246	0.859	15863	0.964
15										6	0.171	1158	0.676	18780	0.923
16										1	**0.167**	287	0.406	15447	0.838
17												25	0.162	7834	0.683
18												1	**0.063**	2007	0.456
19														215	0.254
20														11	**0.184**
	1	1	2	6	21	73	0.936	317	0.898	1662	0.862	10512	0.861	77232	0.864

Table 9. Numbers of 4-Graphs with Distinct Adjacency Spectrum and *Resolution* by Adjacency Spectrum for $n \leq 10$

n	1	2	3	4	5	6		7		8		9		10	
m															
0	1														
1		1													
2			1												
3			1	2											
4				2	3										
5				1	5	5									
6				1	5	12	9	1							
7					4	17	29	1	18	1					
8					2	18	54	0.964	69	0.945	30	**0.857**			
9					1	14	76	**0.962**	175	0.962	174	0.941	73	0.973	
10					1	8	77	0.975	310	0.951	526	0.918	430	**0.905**	
11						3	58	0.983	414	0.963	1195	0.935	1669	0.931	
12						1	31	1	411	0.963	2008	0.929	4481	0.919	
13							9	1	285	0.956	2571	0.929	9388	0.924	
14							2	1	133	**0.933**	2455	0.938	15086	0.916	
15									35	1	1636	0.954	18832	0.926	
16									6	1	656	0.928	17061	0.925	
17											147	0.955	10767	0.938	
18											16	1	4179	0.950	
19													801	0.948	
20													57	0.966	
	1	1	2	6	21	78	345	0.977	1856	0.962	11414	0.935	82824	0.926	

Table 10. Numbers of 4-Graphs with Distinct Distance Spectrum and *Resolution* by Distance Spectrum for $n \leq 10$

n	1	2	3	4	5	6	7		8		9		10	
m														
0	1													
1		1												
2			1											
3			1	2										
4				2	3									
5				1	5	5								
6				1	5	12	9							
7					4	17	29	18	1					
8					2	18	56	73	1	35	1			
9					1	14	79	182	1	185	1	75	1	
10					1	8	79	325	0.997	572	0.998	475	1	
11						3	59	430	1	1276	0.998	1790	0.999	
12						1	31	426	0.998	2161	1	4864	0.998	
13							9	298	1	2764	0.999	10158	0.999	
14							2	134	1	2610	0.998	16447	0.999	
15								35	1	1709	0.997	20323	0.999	
16								6	1	699	0.989	18396	0.998	
17										152	**0.987**	11444	0.997	
18										16	1	4376	0.995	
19												833	**0.986**	
20												59	1	
	1	1	2	6	21	78	353	1927	0.999	12179	0.998	89240	0.998	

bond angles and dihedral angles) will be available to such hypothetical hydrocarbon molecules. Graphs **13** and **14** correspond to substituted tetracyclooctanes of very unusual geometry, more specifically, **13** depicts a bridged [3.2.1]-propellane, **14** a doubly annelated bicyclo[1.1.1]pentane.[30] **15** and **16** correspond to pentacyclononanes, the former to a doubly bridged [3.3.1]propellane, the latter to a bridged [3.2.1]propellane. Not surprisingly, not a single compound containing any of the four polycyclic frameworks of Figure 5 is listed in the Beilstein or the CAS Registry file.

Table 11. Numbers of 4-Graphs with Distinct Combination of *J* and Adjacency Spectrum and *Resolution* by This Combination for $n \leq 10$

n	1	2	3	4	5	6	7	8		9		10	
m													
0	1												
1		1											
2			1										
3			1	2									
4				2	3								
5				1	5	5							
6				1	5	12	9						
7					4	17	29	18					
8					2	18	56	73	35	1			
9					1	14	79	182	185	1	75	1	
10					1	8	79	326	573	1	475	1	
11						3	59	430	1278	1	1792	1	
12						1	31	427	2160	0.999	4875	1	
13							9	298	2765	0.999	10161	0.999	
14							2	134	2611	0.998	16454	0.999	
15								35	1702	0.993	20329	0.999	
16								6	691	**0.977**	18404	0.998	
17									151	0.981	11430	0.996	
18									16	1	4363	0.992	
19											830	**0.982**	
20											59	1	
	1	1	2	6	21	78	353	1929	12167	0.997	89247	0.998	

Table 12. Numbers of 4-Graphs with Distinct Combination of *J* and Distance Spectrum and *Resolution* by This Combination for $n \leq 10$

n	1	2	3	4	5	6	7	8		9		10	
m													
0	1												
1		1											
2			1										
3			1	2									
4				2	3								
5				1	5	5							
6				1	5	12	9						
7					4	17	29	18					
8					2	18	56	73	35	1			
9					1	14	79	182	185	1	75	1	
10					1	8	79	326	573	1	475	1	
11						3	59	430	1278	1	1792	1	
12						1	31	427	2161	1	4875	1	
13							9	298	2767	0.999	10162	1	
14							2	134	2614	0.999	16460	0.999	
15								35	1709	0.999	20341	0.999	
16								6	699	0.989	18423	0.999	
17									152	**0.987**	11458	0.998	
18									16	1	4376	0.995	
19											833	**0.986**	
20											59	1	
	1	1	2	6	21	78	353	1929	12189	0.999	89329	0.999	

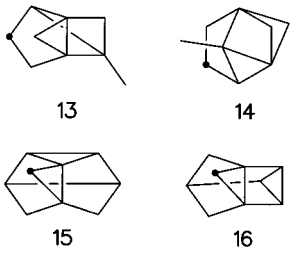

Figure 5. Pair of smallest 4-graphs that are both *J*-equivalent and isospectral (top) and pair of smallest 4-graphs that are both *J*-equivalent and distance-isospectral (and isospectral, bottom). For the meaning of black dots see text.

In the $n = 10$ domain, the smallest pair of *J*-equivalent and isospectral 4-graphs was identified (Table 11) in the $m = 13$ class (tetracyclic decanes), and the smallest pair of *J*-equivalent and distance-isospectral 4-graphs was found to have $m = 14$ edges (pentacyclic decanes, Table 12). Their structures differ from those shown in Figure 5 only in that they bear an additional vertex attached to the one marked with a dot.

From these smallest examples of simultaneously *J*-equivalent and (distance-) isospectral 4-graphs it is concluded that such graphs probably are too complex, too polycyclic for their molecular counterparts to be capable of existence. Other 4-graph pairs of $n = 9$ or 10 being *J*-equivalent and (distance-) isospectral have even higher *m* values, meaning that molecular counterparts would contain even more cycles than those found above, and therefore will tend to be even more strained. This means that it is reasonably safe to use the combination of *J* and (distance) spectrum for identifying distinct molecular substructures and molecular subgraphs, at least in the size range investigated here.

Finally, since NIMSG uses along with *J* only two (distance) eigenvalues rather than the complete (distance) spectrum for discrimination among subgraphs, for the sample of 4-graphs we repeated the described procedure, but using only two adjacency eigenvalues or two distance eigenvalues. By systematic variation it was found that the combinations λ_2 and λ_3 and δ_1 and δ_n (used in two published variants of NIMSG[4]) are not optimal. The most discriminating combinations we were able to find are λ_3 and λ_n (the third and the last adjacency eigenvalues) and δ_2 and δ_{n-1} (the second and second-last distance eigenvalues). As a consequence, NIMSG was now improved accordingly. The results shown in Tables 13 and 14 allow for the estimatation of the "safety" of the new NIMSG versions or the risk of obtaining too few distinct substructures/subgraphs. As was to be expected, the results in Table 13 are somewhat inferior to those in Table 11, those in Table 14 are inferior to those in Table 12. However, resolution losses due to using only two instead of all eigenvalues appear in the high *m* region only, that is for graphs certainly not molecular.

Tree Graphs of $n \leq 20$. For trees (uppermost entry in each column in Tables 1−14) the resolution of the combinations *J* and adjacency spectrum and *J* and distance spectrum is perfect in our graph sample of $n \leq 10$, as expected (recall that the first degeneracy of the adjacency spectrum and of the distance spectrum for trees are known to occur for $n = 8$ and $n = 17$, respectively). To fathom corresponding limits we additionally generated all trees of up to $n = 20$ and searched their *J* values and spectra.[31] The results are given in Table 15. Where differences are found between the resolutions of *J* and the spectra for tree graphs, single index *J* is more discriminating than the complete adjacency spectrum but less discriminating than the distance spectrum. First degeneracies of both *J*/spectrum combinations are encountered for $n = 20$, there are two pairs of *J*-equivalent *and* isospectral such trees (**17/18** and **19/20** in Figure 6), and of these one pair (**19/20**) is even distance isospectral. All these trees are 4-trees, i.e., alkanes, eicosanes. In both pairs the structures differ in a position exchange of ethyl and *gem*-dimethyl substituents, as was discussed earlier.[7a]

Table 13. Numbers of 4-Graphs with Distinct Combination of *J*, λ_3, and λ_n and *Resolution* by This Combination for $n \leq 10$

n	1	2	3	4	5	6	7	8		9		10	
m													
0	1												
1		1											
2			1										
3			1	2									
4				2	3								
5				1	5	5							
6				1	5	12	9						
7					4	17	29	18	1				
8					2	18	56	73	1	35	1		
9					1	14	79	182	1	185	1	75	1
10					1	8	79	326	1	573	1	475	1
11						3	59	430	1	1278	1	1792	1
12						1	31	427	1	2160	0.999	4875	1
13							9	298	1	2765	0.999	10161	0.999
14							2	134	1	2611	0.998	16454	0.999
15								34	**0.971**	1700	0.992	20329	0.999
16								6	1	690	0.976	18404	0.998
17										149	0.968	11428	0.996
18										12	**0.750**	4361	0.991
19												826	0.978
20												55	**0.932**
	1	1	2	6	21	78	353	1928	0.999	12158	0.996	89235	0.998

Table 14. Numbers of 4-Graphs with Distinct Combination of *J*, δ_2, and δ_{n-1} and *Resolution* by This Combination for $n \leq 10$

n	1	2	3	4	5	6	7	8		9		10	
m													
0	1												
1		1											
2			1										
3			1	2									
4				2	3								
5				1	5	5							
6				1	5	12	9						
7					4	17	29	18	1				
8					2	18	56	73	1	35	1		
9					1	14	79	182	1	185	1	75	1
10					1	8	79	326	1	573	1	475	1
11						3	59	430	1	1278	1	1792	1
12						1	31	427	1	2161	1	4875	1
13							9	298	1	2767	0.999	10162	1
14							2	134	1	2614	0.999	16460	0.999
15								34	**0.971**	1707	0.996	20341	0.999
16								6	1	697	0.986	18422	0.999
17										150	0.974	11454	0.998
18										12	**0.750**	4369	0.993
19												830	0.982
20												54	**0.915**
	1	1	2	6	21	78	353	1928	0.999	12179	0.998	89309	0.999

4-Trees of $n \leq 20$. Results for all alkanes C_nH_{2n+2} of up to $n = 20$, generated using MOLGEN 4.0, are given in Table 16. Here as for the general trees *J* is more discriminant than the adjacency spectrum but less than the distance spectrum. Within the alkanes the resolution of *J* is somewhat higher, that of the adjacency spectrum is somewhat lower than within all trees. It was also checked (not shown in the table) that use of λ_3 and λ_n instead of all adjacency eigenvalues and of δ_2 and δ_{n-1} instead of all distance eigenvalues (the NIMSG procedures) does not compromise the complete discrimination among alkanes of up to $n = 19$ (the nonadecanes).

Concluding Remarks. Let us emphasize here once more that structure discrimination by combinations of graph invariants (as done in NIMSG) seems to be a simple but

Table 15. Numbers of Distinct Values, and *Resolution* of *J*, Adjacency Spectrum, Distance Spectrum, Combination of *J* and Adjacency Spectrum, and Combination of *J* and Distance Spectrum, for Trees of $n \leq 20$

		J		adj spectrum		dist spectrum		J/adj sp		J/dist sp	
n	#	#	*res*	#	*res*	#	*res*	#	*res*	#	*res*
1	1	1	*1*	1	*1*	1	*1*	1	*1*	1	*1*
2	1	1	*1*	1	*1*	1	*1*	1	*1*	1	*1*
3	1	1	*1*	1	*1*	1	*1*	1	*1*	1	*1*
4	2	2	*1*	2	*1*	2	*1*	2	*1*	2	*1*
5	3	3	*1*	3	*1*	3	*1*	3	*1*	3	*1*
6	6	6	*1*	6	*1*	6	*1*	6	*1*	6	*1*
7	11	11	*1*	11	*1*	11	*1*	11	*1*	11	*1*
8	23	23	*1*	22	*0.957*	23	*1*	23	*1*	23	*1*
9	47	47	*1*	42	*0.894*	47	*1*	47	*1*	47	*1*
10	106	105	*0.991*	102	*0.962*	106	*1*	106	*1*	106	*1*
11	235	234	*0.996*	204	*0.868*	235	*1*	235	*1*	235	*1*
12	551	537	*0.975*	488	*0.886*	551	*1*	551	*1*	551	*1*
13	1301	1290	*0.992*	1078	*0.829*	1301	*1*	1301	*1*	1301	*1*
14	3159	3026	*0.958*	2723	*0.862*	3159	*1*	3159	*1*	3159	*1*
15	7741	7609	*0.983*	6403	*0.827*	7741	*1*	7741	*1*	7741	*1*
16	19320	18158	*0.940*	16479	*0.853*	19320	*1*	19320	*1*	19320	*1*
17	48629	47480	*0.976*	40313	*0.829*	48628	*0.999*	48629	*1*	48629	*1*
18	123867	114600	*0.925*	106135	*0.857*	123865	*0.999*	123867	*1*	123867	*1*
19	317955	308063	*0.969*	271295	*0.853*	317949	*0.999*	317955	*1*	317955	*1*
20	823065	749284	*0.910*	724455	*0.880*	823051	*0.999*	823063	*0.999*	823064	*0.999*

Table 16. Numbers of Distinct Values, and *Resolution* of *J*, Adjacency Spectrum, Distance Spectrum, Combination of *J* and Adjacency Spectrum, and Combination of *J* and Distance Spectrum, for 4-Trees (Alkanes C_nH_{2n+2}) of $n \leq 20$

		J		adj spectrum		dist spectrum		J/adj sp		J/dist sp	
n	#	#	*res*	#	*res*	#	*res*	#	*res*	#	*res*
1	1	1	*1*	1	*1*	1	*1*	1	*1*	1	*1*
2	1	1	*1*	1	*1*	1	*1*	1	*1*	1	*1*
3	1	1	*1*	1	*1*	1	*1*	1	*1*	1	*1*
4	2	2	*1*	2	*1*	2	*1*	2	*1*	2	*1*
5	3	3	*1*	3	*1*	3	*1*	3	*1*	3	*1*
6	5	5	*1*	5	*1*	5	*1*	5	*1*	5	*1*
7	9	9	*1*	9	*1*	9	*1*	9	*1*	9	*1*
8	18	18	*1*	18	*1*	18	*1*	18	*1*	18	*1*
9	35	35	*1*	30	*0.857*	35	*1*	35	*1*	35	*1*
10	75	75	*1*	73	*0.973*	75	*1*	75	*1*	75	*1*
11	159	159	*1*	136	*0.855*	159	*1*	159	*1*	159	*1*
12	355	349	*0.983*	307	*0.865*	355	*1*	355	*1*	355	*1*
13	802	799	*0.996*	652	*0.813*	802	*1*	802	*1*	802	*1*
14	1858	1808	*0.973*	1580	*0.850*	1858	*1*	1858	*1*	1858	*1*
15	4347	4305	*0.990*	3484	*0.801*	4347	*1*	4347	*1*	4347	*1*
16	10359	9923	*0.958*	8573	*0.828*	10359	*1*	10359	*1*	10359	*1*
17	24894	24516	*0.985*	19786	*0.795*	24893	*0.999*	24894	*1*	24894	*1*
18	60523	57331	*0.947*	50340	*0.832*	60521	*0.999*	60523	*1*	60523	*1*
19	148284	145206	*0.979*	122453	*0.826*	148279	*0.999*	148284	*1*	148284	*1*
20	366319	342886	*0.936*	313498	*0.856*	366308	*0.999*	366317	*0.999*	366318	*0.999*

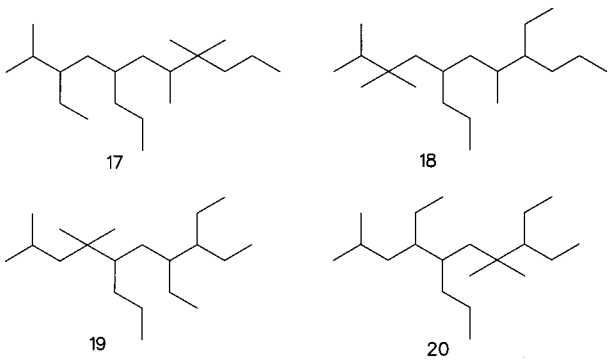

Figure 6. Two pairs of both *J*-equivalent and isospectral eicosanes. The bottom pair is distance-isospectral as well. These are the smallest alkane graphs having these properties.

only approximate solution to a difficult problem. Here we considered graphs corresponding to a superset of saturated hydrocarbons (of rather low carbon number) only, so we cannot say anything about the discrimination of real chemical structures other than saturated hydrocarbons. Most molecular structures, containing multiple bonds and heteroatoms, are to be represented by colored multigraphs. Obviously, there are many more colored multigraphs than simple graphs for each vertex number *n*, so that their discrimination seems to be even more difficult. On the other hand, we carefully included information on multiple bonds and heteroatoms into *J* and the spectra used in NIMSG,[4,32] hopefully raising the discriminating power of the procedure to a level sufficient for practical purposes. Further, in mathematical graph theory experience is that the graph isomorphism problem is more difficult for simple graphs than for colored multigraphs, the former lacking distinguishing features. To test this point would require one to have a comprehensive sample of molecular colored multigraphs, which obviously is not at hand for any *n*.

After proving that "almost all trees are cospectral", Schwenk raised the question whether the same is true for almost all graphs.[12] He did, however, not answer this question, nor did he give a conjecture, due to considerable differences in the mathematical properties of trees on one side and (general) graphs on the other. We here obtained at least some experimental information relevant to this issue. In the adjacency spectrum column in Table 15, resolution values oscillate and only slowly decrease for increasing *n*, so that one would probably not have predicted Schwenk's result. In comparison, the resolution values in Table 3 rapidly and monotonically decrease for increasing *n*, so that *a forteriori* it may seem probable that they drop below 0.5 for some higher *n* (At resolution 0.5 each graph on average has a nondistinguished mate.).

It is tempting to ask similar questions with respect to the other graph invariant (combinations) considered here. The resolution of *J* for general graphs rapidly decreases for increasing *n* (Table 2), dropping to 0.31 for $n = 10$ already, so that from this experimental point of view almost all graphs are *J*-equivalent (i.e. have a *J*-equivalent mate). The situation is less clear for *J* and the trees. Though *J* is still one of the best-discriminating simple invariants for trees (as we saw it is even better than the adjacency spectrum in this respect, Table 15), our data do not exclude the possibility that almost all trees are *J*-equivalent. This may seem paradoxic, but it is not a contradiction.

On the limited data obtained here for the distance spectrum and the *J*/spectrum combinations we do not want to speculate. Their resolutions also drop for increasing *n*, but slowly and not always monotonically, so that it seems possible but by no means clear that statements similar to the above are true for them.

ACKNOWLEDGMENT

Help with the handling of very large files and reformatting lists of connection tables by Dipl.-Math. J. Braun and Dipl.-Math R. Gugisch is gratefully acknowledged. We thank Dr. M. Grohe for ref 16.

REFERENCES AND NOTES

(1) (a) Bonchev, D. The Overall Wiener Index − A New Tool for Characterization of Molecular Topology. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 582−592. (b) Bertz, S. H. Complexity of Molecules and their Synthesis. In *Complexity in Chemistry*; Bonchev, D., Rouvray, D. H., Eds.; Mathematical Chemistry Series, Vol. 7, in press. (c) Varmuza, K.; Scsibrany, H. Substructure Isomorphism Matrix. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 308−313. (d) Poroikov, V. V.; Filimonov, D. A.; Borodina, Yu. V.; Lagunin, A. A.; Kos, A. Robustness of Biological Activity Spectra Predicting by Computer Program PASS for Noncongeneric Sets of Chemical Compounds. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 1349−1355.

(2) (a) Blinov, K. A.; Elyashberg, M. E.; Molodtsov, S. G.; Williams, A. J.; Martirosian, E. R. An Expert System for Automated Structure Elucidation Utilizing ^{1}H-^{1}H, ^{13}C-^{1}H and ^{15}N-^{1}H 2D NMR Correlations. *Fresenius J. Anal. Chem.* **2001**, *369*, 709−714. (b) Varmuza, K.; Penchev, P. N.; Scsibrany, H. Maximum Common Substructures of Organic Compounds Exhibiting Similar Infrared Spectra. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 420−427. (c) Seebass, B.; Pretsch, E. Automated Compatibility Tests of the Molecular Formulas or Structures of Organic Compounds with Their Mass Spectra. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 713−717.

(3) Rücker, G.; Rücker, C. Automatic Enumeration of All Connected Subgraphs. *MATCH − Commun. Math. Comput. Chem.* **2000**, *41*, 145−149.

(4) Rücker, G.; Rücker, C. On Finding Nonisomorphic Connected Subgraphs and Distinct Molecular Substructures. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 314−320; 865.

(5) Rücker, G.; Rücker, C. Substructure, Subgraph, and Walk Counts as Measures of the Complexity of Graphs and Molecules. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 1457−1462.

(6) Balaban, A. T. Higly Discriminating Distance-Based Topological Index. *Chem. Phys. Lett.* **1982**, *89*, 399−404.

(7) (a) Balaban, A. T.; Quintas, L. V. The Smallest Graphs, Trees, and 4-Trees with Degenerate Topological Index J. *MATCH − Commun. Math. Comput. Chem.* **1983**, *14*, 213−233. (b) Razinger, M.; Chrétien, J. R.; Dubois, J. E. Structural Selectivity of Topological Indexes in Alkane Series. *J. Chem. Inf. Comput. Sci.* **1985**, *25*, 23−27.

(8) (a) Bonchev, D.; Mekenyan, O.; Trinajstić, N. Isomer Discrimination by Topological Information Approach. *J. Comput. Chem.* **1981**, *2*, 127−148. (b) Balasubramanian, K.; Basak, S. C. Characterization of Isospectral Graphs Using Graph Invariants and Derived Orthogonal Parameters. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 367−373.

(9) Randić, M.; Vracko, M.; Nović, M. Eigenvalues as Molecular Descriptors. In Diudea, M. V., Ed.; *QSPR/QSAR Studies by Molecular Descriptors*; Nova Science Publ.: Huntington, NY, 2001.

(10) Balaban, A. T.; Harary, F. The Characteristic Polynomial Does Not Uniquely Determine the Topology of a Molecule. *J. Chem. Doc.* **1971**, *11*, 258−259.

(11) Cvetković, D.; Rowlinson, P.; Simić, S. *Eigenspaces of Graphs*; Cambridge University Press: Cambridge, 1997.

(12) Schwenk, A. J. Almost All Trees are Cospectral. In *New Directions in the Theory of Graphs;* Harary, F., Ed.; Academic Press: New York, 1973; pp 275−307.

(13) Mihalić, Z.; Veljan, D.; Amić, D.; Nikolić, S.; Plavsić, D.; Trinajstić, N. The Distance Matrix in Chemistry. *J. Math. Chem.* **1992**, *11*, 223−258.

(14) Cvetković, D. M.; Doob, M.; Gutman, I.; Torgasev, A. *Recent Results in the Theory of Graph Spectra*; Elsevier: Amsterdam, 1988; p 128.

(15) McKay, B. D. On the Spectral Characterisation of Trees. *Ars Combinatorica* **1977**, *3*, 219−232.

(16) Cai, J.-Y.; Fürer, M.; Immerman, N. An Optimal Lower Bound on the Number of Variables for Graph Identification. *Combinatorica* **1992**, *4*, 389−410.

(17) Shrikhande, S. S. On a Characterization of the Triangular Association Scheme. *Ann. Math. Statist.* **1959**, *30*, 39−47.

(18) Weisfeiler, B. *On Construction and Identification of Graphs*; Lecture Notes in Mathematics No. 558, Springer: Berlin, 1976.

(19) Mathon, R. *Proc. 9th S.-E. Conf. Combinatorics, Graph Theory and Computing* **1978**, 499.

(20) Read, R. C.; Wilson, R. J. *An Atlas of Graphs*; Clarendon Press: Oxford: 1998; pp 4 and 7.

(21) Recall that cubane, cuneane, and octabisvalene are pentacyclic octanes ($n = 8$, $m = 12$, $c = 5$). A heptacyclic octane ($n = 8$, $m = 14$, $c = 7$) would require two additional bonds in a cubane skeleton, for example.

(22) Kerber, A.; Laue, R.; Grüner, T.; Meringer, M. MOLGEN 4.0. *MATCH − Commun. Math. Comput. Chem.* **1998**, *37*, 205−208.

(23) *J* and eigenvalues were calculated as double precision numbers, for comparisons eight decimal places and seven decimal places were used for *J* and eigenvalues, respectively, for reasons detailed in ref 4.

(24) We note in passing that having the necessary software we also determined the triplet of smallest isospectral 4-graphs ($n = 8$, $m = 10$, corresponding to 2-methyltricyclo[3.2.0.0^{1,6}]heptane, 3,4-dimethyltricyclo[3.1.0.0^{2,6}]hexane, and [4.1.1]propellane) and the quadruplet of smallest isospectral graphs ($n = 8$, $m = 11$, corresponding to 7-methyltetracyclo[2.2.1.0.^{1,3}0^{1,5}]heptane, 6-methyltetracyclo[3.1.1.0.^{1,3}0^{3,5}]heptane, tetracyclo[2.2.2.0.^{1,3}0^{1,4}]octane, and 1-methyltetracyclo[3.2.0.0.^{1,3}0^{2,7}]-heptane). Recently an isospectral triplet of $n = 9$, $m = 16$ and an isospectral quadruplet of $n = 9$, $m = 19$ were published.[26b] Further we identified the pair of smallest distance-isospectral connected simple graphs, which have $n = 7$, $m = 10$, corresponding to tetracyclo[3.1.1.0.^{1,3}0^{3,5}]heptane and tetracyclo-[2.2.1.0.^{1,3}0^{1,5}]heptane.

(25) In fact, this was not so surprising, after these graphs could not be pairwise distinguished using *J* and the distance eigenvalues.

(26) (a) Hosoya, H.; Nagashima, U.; Hyugaji, S. Topological Twin Graphs. Smallest Pair of Isospectral Polyhedral Graphs with Eight vertices. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 428−431. (b) Hosoya, H.; Ohta, K.; Satomi, M. Topological Twin Graphs II. Isospectral Polyhedral Graphs With Nine and Ten vertices. *MATCH − Commun. Math. Comput. Chem.* **2001**, *44*, 183−200.

(27) Despite having many graph invariants identical, such pairs of graphs can be differentiated by their chemical names, easily obtained using program POLCYC, e.g. **1** corresponds to heptacyclo[3.3.0.0.^{1,3}-0.^{1,4}0.^{2,6}0.^{2,7}0^{6,8}]octane, **2** to heptacyclo[4.2.0.0.^{1,3}0.^{1,7}0.^{2,4}0.^{2,5}0^{5,8}]-octane.

(28) If index values are compared in a sample of several or all edge numbers *m* within a constant vertex number *n*, i.e., without prior sorting by *m*, then additional degeneracies will occur. For instance, cyclooctane ($n = 8$, $m = 8$) and cubane ($n = 8$, $m = 12$) share the *J* value 2.000. Since NIMSG always sorts by *n* and *m*, we are not interested in such degeneracies here. It is well-known that graphs of *different n* can share the same *J* value, e.g. cyclohexane ($n = 6$, $m = 6$) also has $J = 2.000$.

(29) Initially we were concerned that *J* and the distance spectrum, both being derived from the distance matrix, might tend to exhibit degeneracies for the same pairs of graphs. Fortunately, as foreseen already from the results shown in Figures 1 and 2, such concerns did not materialize to a large extent. However, the moderate improvement in the resolution of the distance spectrum on addition of *J* compared to the large improvement in the resolution of the adjacency spectrum on addition of *J* (Tables 15/16) may be interpreted to be partially due to such an effect.

(30) The lower homologues of **13** and **14** lacking the methyl group are neither *J*-equivalent nor isospectral.

(31) Grüner, T. Program GRADPART; Diploma Thesis, University of Bayreuth, 1995.

(32) Balaban, A. T. Topological Index J for Heteroatom-Containing Molecules Taking into Account Periodicities of Element Properties. *MATCH − Commun. Math. Comput. Chem.* **1986**, *21*, 115−122.

CI010121Y