

KnowItAll ソフトウェアのトレーニング

データマイニング&解析

データマイニング&解析

オーバーラップ密度ヒートマップ: スペクトル、クロマトグラム、および他のグラフィカルデータを分析するための技術

目的

この演習では、オーバーラップ密度ヒートマップを使用したデータマイニングと可視化の手法を紹介します。

目標

この演習では、以下の内容を学ぶことができます:

- ▶ オーバーラップ密度ヒートマップの表示と操作方法

背景

Wiley の特許を取得したオーバーラップ密度ヒートマップは、大量のスペクトル、クロマトグラム、および他のグラフィカルデータの類似性と相違点を評価するのに役立ちます。

この技術により、スペクトルやクロマトグラムなどのオーバーラップしたオブジェクトの共通の特徴を、重なる高い部分から低い部分まで色分けして視覚的に表示することができます。

このレッスンで使用されるトレーニングファイルは、以下の場所に保存されています

C:\Users\Public\Documents\Wiley\KnowItAll\Samples\IR
フォルダに移動します。

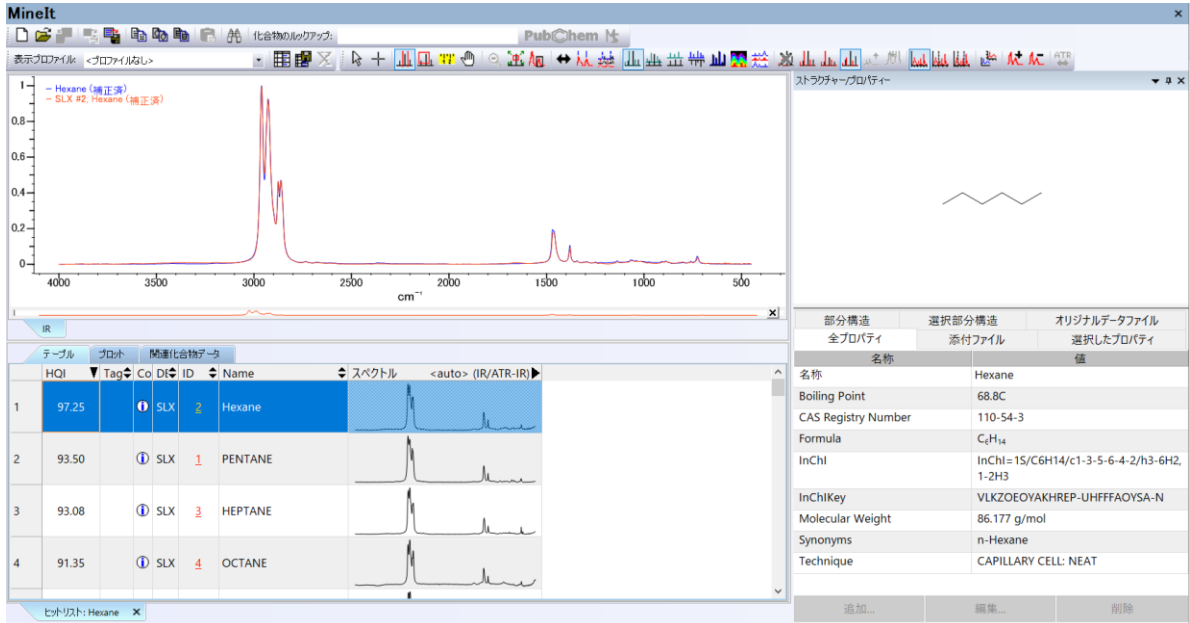
- Hexane.jdx

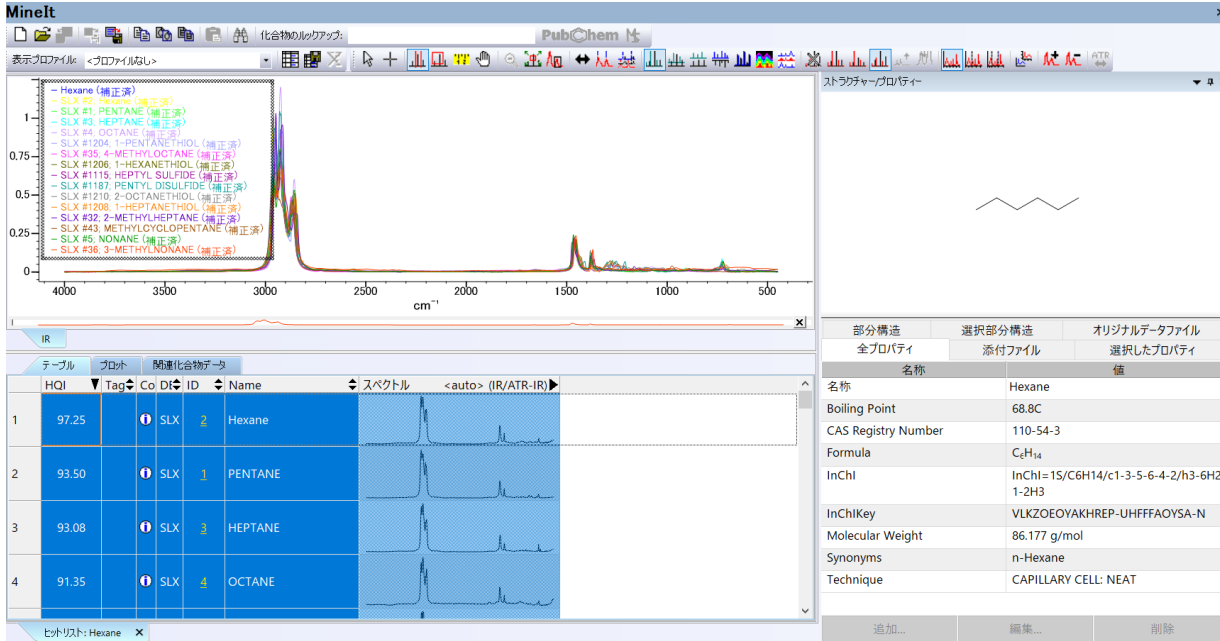
KnowItAll 使用アプリケーション

- Minelt™


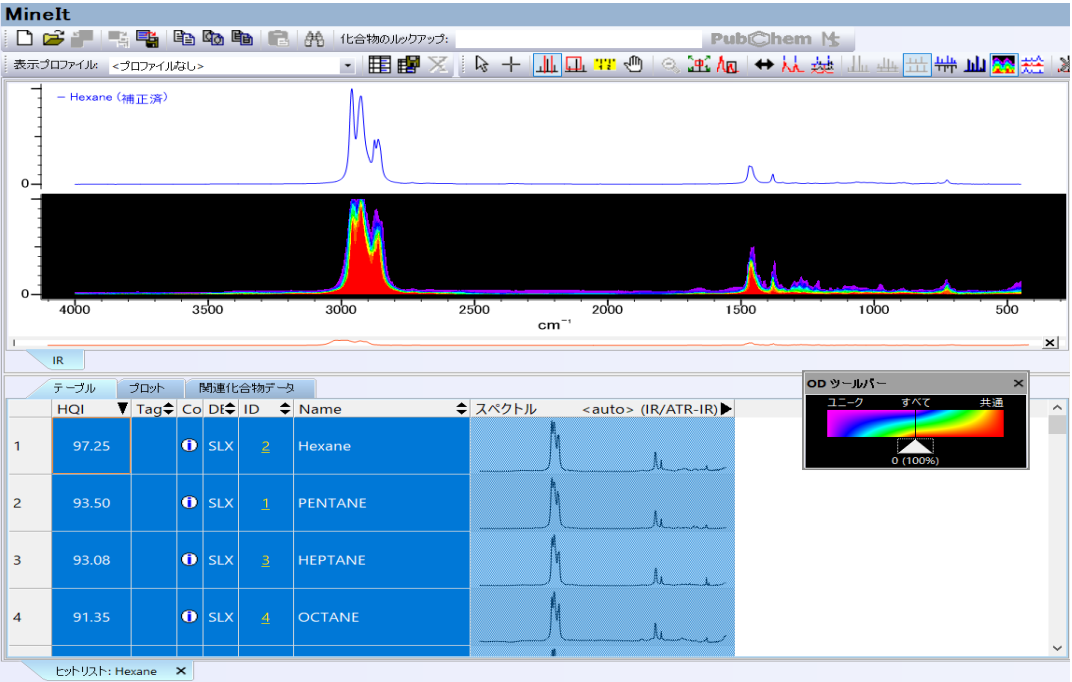
Minelt アプリケーションでスペクトルのヒットリストを開く

	アクション	結果
1	<p>SearchIt アプリケーションで Search Databases の下に ある User-Select をクリックします。</p> <p>もし、Selected for Searching パネルにデータベースが表 示されている場合は、すべて削除をクリックします。</p> <p>IR - Sadtler Standards (Selected Subset) - Wiley (DB Code SLX)を Selected for Searching パネルに追加しま す。</p>	
2	<p>Search Categories パネルで Spectrum をクリックしま す。</p> <p>C:\Users\Public\Documents\Wiley\KnowItAll\Samples\IR フォルダに移動します。</p> <p>Hexane.jdx を開きます。</p>	

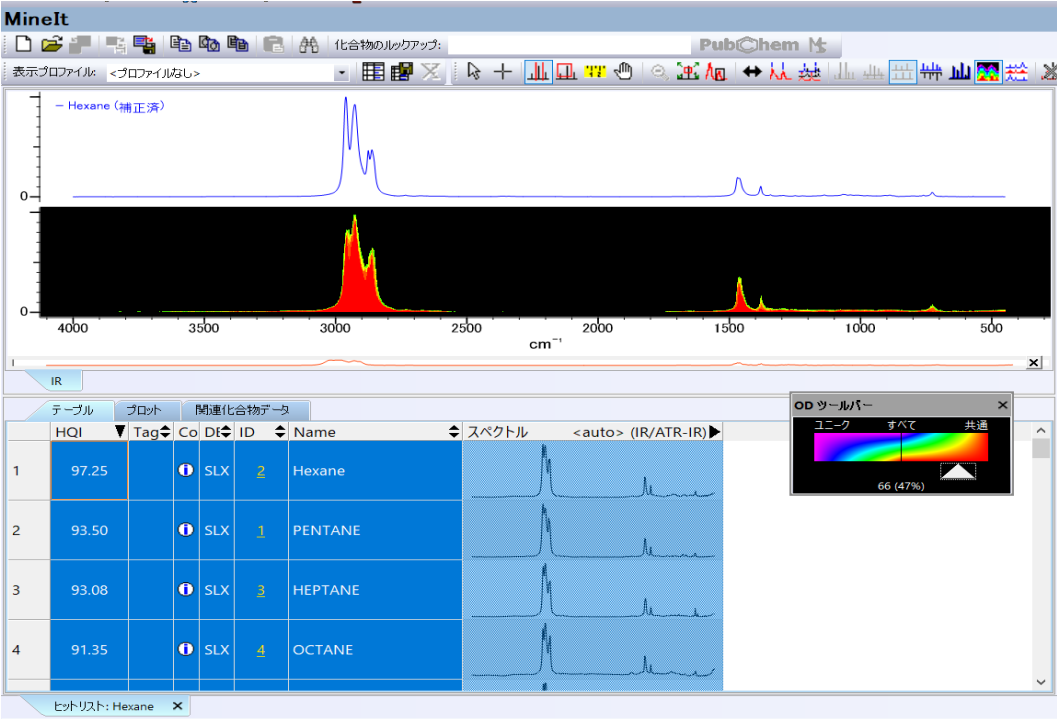
	アクション	結果
3	「Search」をクリックします。	<p>Minelt アプリケーションにて結果が表示されます。</p>  <p>The screenshot displays the Minelt application interface. At the top, there is a toolbar with various icons. Below the toolbar is a plot area showing an IR spectrum with two traces: a blue trace for 'Hexane (修正済)' and a red trace for 'SLX #2: Hexane (修正済)'. The x-axis is labeled 'cm⁻¹' and ranges from 4000 to 500. Below the plot is a table with columns: HQI, Tag, Co, DI, ID, Name, and スペクトル. The table lists four entries: 1. Hexane (HQI: 97.25, Co: SLX, ID: 2), 2. PENTANE (HQI: 93.50, Co: SLX, ID: 1), 3. HEPTANE (HQI: 93.08, Co: SLX, ID: 3), and 4. OCTANE (HQI: 91.35, Co: SLX, ID: 4). To the right of the table is a chemical structure diagram of Hexane, represented as a zigzag line. Below the structure is a table with columns: 部分構造, 選択部分構造, オリジナルデータファイル, 全プロパティ, 添付ファイル, and 選択したプロパティ. The table lists various properties for Hexane, including Boiling Point (68.8C), CAS Registry Number (110-54-3), Formula (C₆H₁₄), InChI, InChIKey, Molecular Weight (86.177 g/mol), Synonyms (n-Hexane), and Technique (CAPILLARY CELL: NEAT).</p>

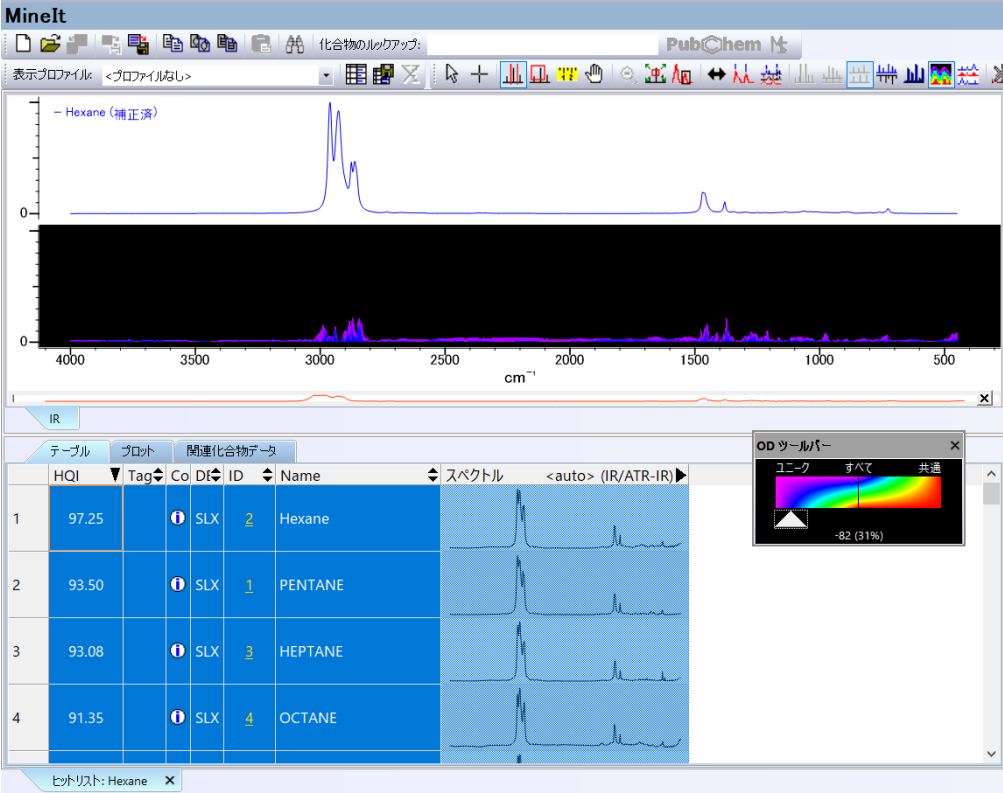
	アクション	結果																																				
4	<p>ヒットリストで最初の 15 のエントリを選択します。</p> <p>注記: 最初のレコードを選択し、Shift キーを押しながら 15 番目のエントリまでスクロールし、選択します。</p>	<p>選択したすべてのヒットがスペクトルパネルに表示されます。</p>  <table border="1" data-bbox="1501 706 1900 974"> <thead> <tr> <th>部分構造</th> <th>選択部分構造</th> <th>オリジナルデータファイル</th> </tr> <tr> <th>全プロパティ</th> <th>添付ファイル</th> <th>選択したプロパティ</th> </tr> <tr> <th>名称</th> <th>名称</th> <th>値</th> </tr> </thead> <tbody> <tr> <td>名称</td> <td>Hexane</td> <td></td> </tr> <tr> <td>Boiling Point</td> <td>68.8C</td> <td></td> </tr> <tr> <td>CAS Registry Number</td> <td>110-54-3</td> <td></td> </tr> <tr> <td>Formula</td> <td>C₆H₁₄</td> <td></td> </tr> <tr> <td>InChI</td> <td>InChI=1S/C6H14/c1-3-5-6-4-2/h3-6H2,1-2H3</td> <td></td> </tr> <tr> <td>InChIKey</td> <td>VLKZOEYAKHREP-UHFFFAOYSA-N</td> <td></td> </tr> <tr> <td>Molecular Weight</td> <td>86.177 g/mol</td> <td></td> </tr> <tr> <td>Synonyms</td> <td>n-Hexane</td> <td></td> </tr> <tr> <td>Technique</td> <td>CAPILLARY CELL: NEAT</td> <td></td> </tr> </tbody> </table>	部分構造	選択部分構造	オリジナルデータファイル	全プロパティ	添付ファイル	選択したプロパティ	名称	名称	値	名称	Hexane		Boiling Point	68.8C		CAS Registry Number	110-54-3		Formula	C ₆ H ₁₄		InChI	InChI=1S/C6H14/c1-3-5-6-4-2/h3-6H2,1-2H3		InChIKey	VLKZOEYAKHREP-UHFFFAOYSA-N		Molecular Weight	86.177 g/mol		Synonyms	n-Hexane		Technique	CAPILLARY CELL: NEAT	
部分構造	選択部分構造	オリジナルデータファイル																																				
全プロパティ	添付ファイル	選択したプロパティ																																				
名称	名称	値																																				
名称	Hexane																																					
Boiling Point	68.8C																																					
CAS Registry Number	110-54-3																																					
Formula	C ₆ H ₁₄																																					
InChI	InChI=1S/C6H14/c1-3-5-6-4-2/h3-6H2,1-2H3																																					
InChIKey	VLKZOEYAKHREP-UHFFFAOYSA-N																																					
Molecular Weight	86.177 g/mol																																					
Synonyms	n-Hexane																																					
Technique	CAPILLARY CELL: NEAT																																					

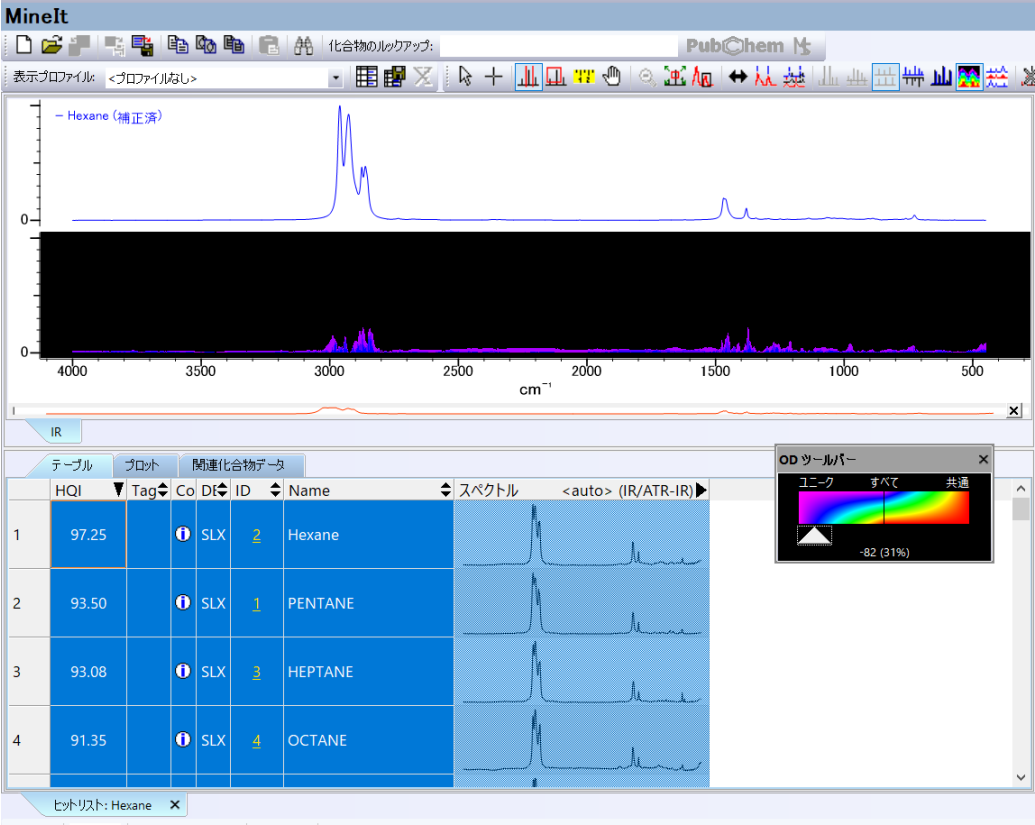
スペクトル表示をオーバーラップ密度ヒートマップに変更

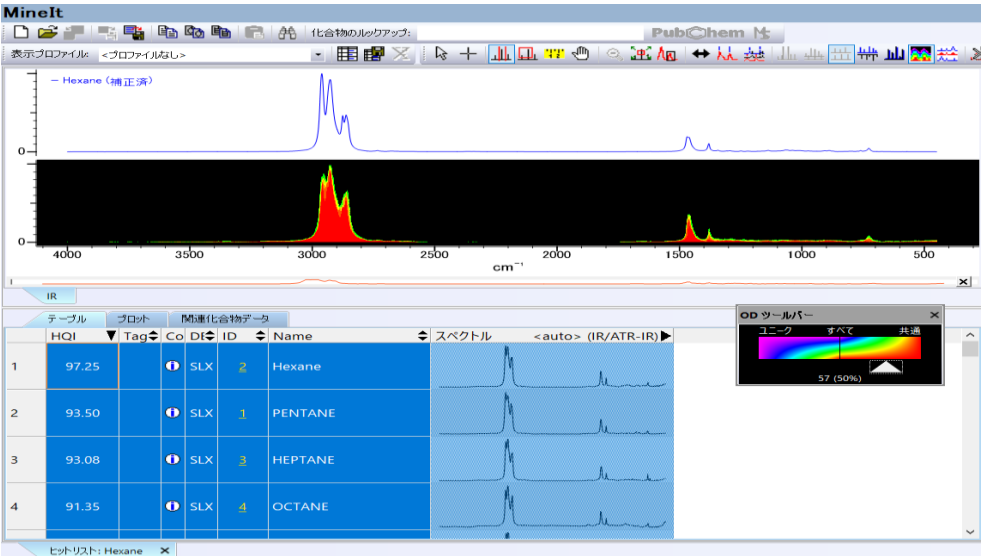
アクション	結果
1 スペクトルツールバーで、オーバーラップ密度 (OD) ヒートマップツール  をクリックします。	<p>すると、スペクトル表示が変わり、オーバーラップ密度 (OD) ツールバーが表示されます。</p>  <p>デフォルトのオーバーラップ密度ヒートマップでは、すべてのオーバーラップレベルが表示されます。高いオーバーラップレベルは赤で表示され、低いレベルは紫で表示されます。OD ツールバーを使うと、スライダーを操作して表示するオーバーラップの量を調整できます。デフォルトのオーバーラップ密度ヒートマップでは、スライダーは OD レベル=0 に設定されており、全てのオーバーラップ密度レベルに対応する色が表示されます。</p>

オーバーラップ密度ヒートマップを操作

	アクション	結果																																			
1	OD ツールバー のスライダーを右に動かしてください。	<p>スライダーを右に動かすと、共通のオーバーラップ領域が表示されます。OD レベルが 100 に近づくと、最も共通した領域のみが表示されます。</p>  <p>The screenshot displays the MineIt interface. At the top, there's a toolbar and a 'PubChem' logo. Below that, a plot area shows IR spectra for 'Hexane (補正済)'. The x-axis is labeled 'cm⁻¹' and ranges from 4000 to 500. A density heatmap is overlaid on the spectra. Below the plot is a table with columns: HQI, Tag, Co, DI, ID, Name, and スペクトル. The table lists four entries: Hexane (HQI: 97.25, ID: 2), PENTANE (HQI: 93.50, ID: 1), HEPTANE (HQI: 93.08, ID: 3), and OCTANE (HQI: 91.35, ID: 4). To the right of the table is an 'OD ツールバー' (OD Tool) with a color scale and a slider. The slider is positioned at 66 (47%), and the selected mode is '共通' (Common). Below the table, it says 'ヒットリスト: Hexane'.</p> <table border="1"><thead><tr><th>HQI</th><th>Tag</th><th>Co</th><th>DI</th><th>ID</th><th>Name</th><th>スペクトル</th></tr></thead><tbody><tr><td>97.25</td><td></td><td>1</td><td>SLX</td><td>2</td><td>Hexane</td><td></td></tr><tr><td>93.50</td><td></td><td>1</td><td>SLX</td><td>1</td><td>PENTANE</td><td></td></tr><tr><td>93.08</td><td></td><td>1</td><td>SLX</td><td>3</td><td>HEPTANE</td><td></td></tr><tr><td>91.35</td><td></td><td>1</td><td>SLX</td><td>4</td><td>OCTANE</td><td></td></tr></tbody></table>	HQI	Tag	Co	DI	ID	Name	スペクトル	97.25		1	SLX	2	Hexane		93.50		1	SLX	1	PENTANE		93.08		1	SLX	3	HEPTANE		91.35		1	SLX	4	OCTANE	
HQI	Tag	Co	DI	ID	Name	スペクトル																															
97.25		1	SLX	2	Hexane																																
93.50		1	SLX	1	PENTANE																																
93.08		1	SLX	3	HEPTANE																																
91.35		1	SLX	4	OCTANE																																

	アクション	結果																																							
2	スライダーを左に動かしてください。	<p>OD レベルが-100 に近づくと、最もオーバーラップが少ないユニークな領域のみが表示されます。</p>  <p>The screenshot shows the MineIt interface with an IR spectrum plot and a table of results. The table lists four compounds: Hexane (HQI 97.25), PENTANE (HQI 93.50), HEPTANE (HQI 93.08), and OCTANE (HQI 91.35). An OD slider is visible on the right side of the table, set to -82 (31%).</p> <table border="1" data-bbox="699 865 1381 1174"> <thead> <tr> <th>テーブル</th> <th>プロット</th> <th>関連化合物データ</th> </tr> <tr> <th>HQI</th> <th>Tag</th> <th>Co</th> <th>DI</th> <th>ID</th> <th>Name</th> <th>スペクトル</th> <th><auto> (IR/ATR-IR)</th> </tr> </thead> <tbody> <tr> <td>1</td> <td>97.25</td> <td>SLX</td> <td>2</td> <td>Hexane</td> <td>[Spectrum]</td> <td>[Spectrum]</td> </tr> <tr> <td>2</td> <td>93.50</td> <td>SLX</td> <td>1</td> <td>PENTANE</td> <td>[Spectrum]</td> <td>[Spectrum]</td> </tr> <tr> <td>3</td> <td>93.08</td> <td>SLX</td> <td>3</td> <td>HEPTANE</td> <td>[Spectrum]</td> <td>[Spectrum]</td> </tr> <tr> <td>4</td> <td>91.35</td> <td>SLX</td> <td>4</td> <td>OCTANE</td> <td>[Spectrum]</td> <td>[Spectrum]</td> </tr> </tbody> </table> <p>OD ツールバー ユニーク すべて 共通 -82 (31%)</p> <p>OD レベルを変更するたびに、括弧内には第二の値が表示されます。この値は、OD レベル= 0 のときの曲線下の面積の割合である%AUC です。</p>	テーブル	プロット	関連化合物データ	HQI	Tag	Co	DI	ID	Name	スペクトル	<auto> (IR/ATR-IR)	1	97.25	SLX	2	Hexane	[Spectrum]	[Spectrum]	2	93.50	SLX	1	PENTANE	[Spectrum]	[Spectrum]	3	93.08	SLX	3	HEPTANE	[Spectrum]	[Spectrum]	4	91.35	SLX	4	OCTANE	[Spectrum]	[Spectrum]
テーブル	プロット	関連化合物データ																																							
HQI	Tag	Co	DI	ID	Name	スペクトル	<auto> (IR/ATR-IR)																																		
1	97.25	SLX	2	Hexane	[Spectrum]	[Spectrum]																																			
2	93.50	SLX	1	PENTANE	[Spectrum]	[Spectrum]																																			
3	93.08	SLX	3	HEPTANE	[Spectrum]	[Spectrum]																																			
4	91.35	SLX	4	OCTANE	[Spectrum]	[Spectrum]																																			

	アクション	結果																																											
3	スライダーを動かし、%AUC が約 50% になる地点まで移動してください。	<p>最も共通の領域の 50% を表示するためには、共通側の OD レベル (ODC) が 62 である必要があります。</p>  <p>The screenshot shows the MineIt interface with the following table of results:</p> <table border="1"> <thead> <tr> <th>Table</th> <th>Plot</th> <th>Related Compound Data</th> </tr> <tr> <th>HQI</th> <th>Tag</th> <th>Co</th> <th>DI</th> <th>ID</th> <th>Name</th> <th>Spectrum</th> <th><auto> (IR/ATR-IR)</th> </tr> </thead> <tbody> <tr> <td>1</td> <td>97.25</td> <td></td> <td>1</td> <td>2</td> <td>Hexane</td> <td>[Spectrum]</td> <td>[Spectrum]</td> </tr> <tr> <td>2</td> <td>93.50</td> <td></td> <td>1</td> <td>1</td> <td>PENTANE</td> <td>[Spectrum]</td> <td>[Spectrum]</td> </tr> <tr> <td>3</td> <td>93.08</td> <td></td> <td>1</td> <td>3</td> <td>HEPTANE</td> <td>[Spectrum]</td> <td>[Spectrum]</td> </tr> <tr> <td>4</td> <td>91.35</td> <td></td> <td>1</td> <td>4</td> <td>OCTANE</td> <td>[Spectrum]</td> <td>[Spectrum]</td> </tr> </tbody> </table> <p>Hit list: Hexane</p> <p>このことを別の言い方で表現すると、$ODC_{50} = 62$ と言えます。また、$\%AUC_{42} = 50$ とも言えます。</p>	Table	Plot	Related Compound Data	HQI	Tag	Co	DI	ID	Name	Spectrum	<auto> (IR/ATR-IR)	1	97.25		1	2	Hexane	[Spectrum]	[Spectrum]	2	93.50		1	1	PENTANE	[Spectrum]	[Spectrum]	3	93.08		1	3	HEPTANE	[Spectrum]	[Spectrum]	4	91.35		1	4	OCTANE	[Spectrum]	[Spectrum]
Table	Plot	Related Compound Data																																											
HQI	Tag	Co	DI	ID	Name	Spectrum	<auto> (IR/ATR-IR)																																						
1	97.25		1	2	Hexane	[Spectrum]	[Spectrum]																																						
2	93.50		1	1	PENTANE	[Spectrum]	[Spectrum]																																						
3	93.08		1	3	HEPTANE	[Spectrum]	[Spectrum]																																						
4	91.35		1	4	OCTANE	[Spectrum]	[Spectrum]																																						

アクション	結果
<p>4 次に、スライダーを左に動かして、%AUC が約 50%になる位置に移動してください。</p>	<p>最もユニークな領域の 50%を表示するためには、ユニーク側の OD レベル (ODU) が-46 が必要です。</p>  <p>このことを別の言い方で表現すると、$ODC_{50} = -64$ と言えます。また、$\%AUC_{-64} = 50$ とも言えます。</p>
<p>ヒント</p>	<p>異なる領域を調べるためには、水平ズームツールを使用します。まず、水平ズームツールを選択し、スペクトル上の領域をクリックしてドラッグしてズームインします。全体のスペクトルを表示するためには、スペクトル全体表示ツールを使用してズームアウトします。</p>

データマイニング&解析

オーバーラップ密度コンセンサスペクトルの作成と使用方法

目的

演習では、KnowItAll 情報システムでオーバーラップ密度コンセンサスペクトルを作成し、利用する方法をご紹介します。

目標

この演習では、以下の内容を学ぶことができます:

- オーバーラップ密度コンセンサスペクトルの表示と操作方法
- オーバーラップ密度コンセンサスペクトルを表示し、操作する方法

背景

Wiley の特許取得済みのオーバーラップ密度ヒートマップ技術は、大量のスペクトル、クロマトグラム、およびその他のグラフィカルデータの類似性や相違点を視覚的にデータマイニングや分析するために有用です。

特定の OD レベルで、最も高いオーバーラップ領域の輪郭を追跡することにより、OD ヒートマップ内の各スペクトル x 値において最大のスペクトル y 値を使って、数学的に複合スペクトルを作り出すことができます。このオーバーラップ密度コンセンサスペクトルは、類似したスペクトルを検索するためにスペクトル検索に利用できます。また、将来の利用のためにデータベースに保存することも可能です。

このレッスンで使用されるトレーニングファイルは、以下の場所に保存されています


C:\Users\Public\Documents\Wiley\KnowItAll\Samples\IR フォルダに移動します。

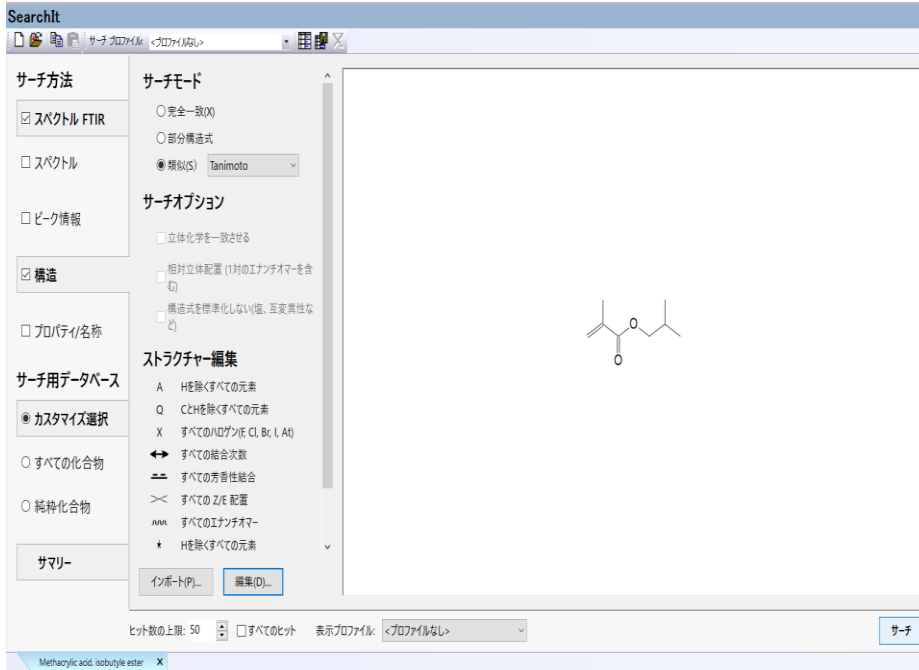
- Propiophenone Query.dsf

KnowItAll 使用アプリケーション

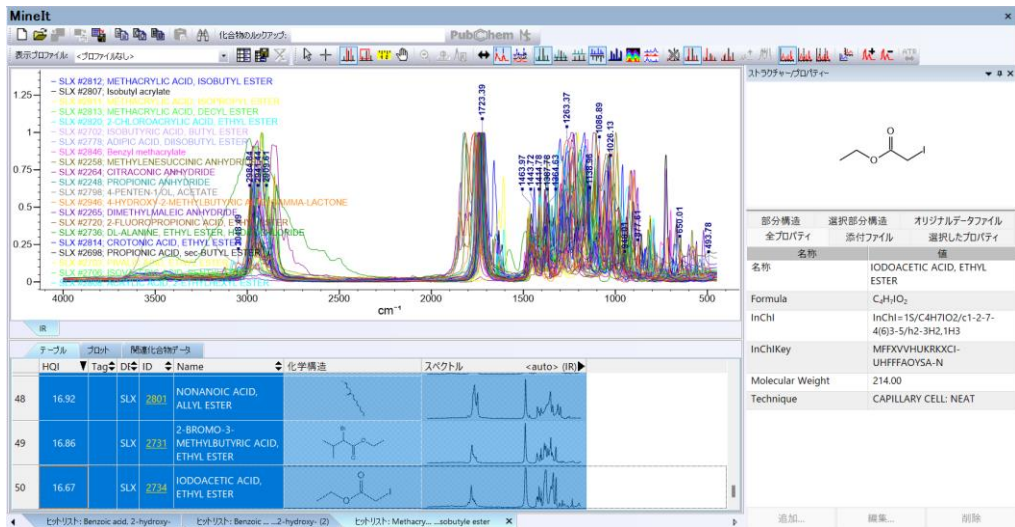
- SearchIt™
- Minelt™


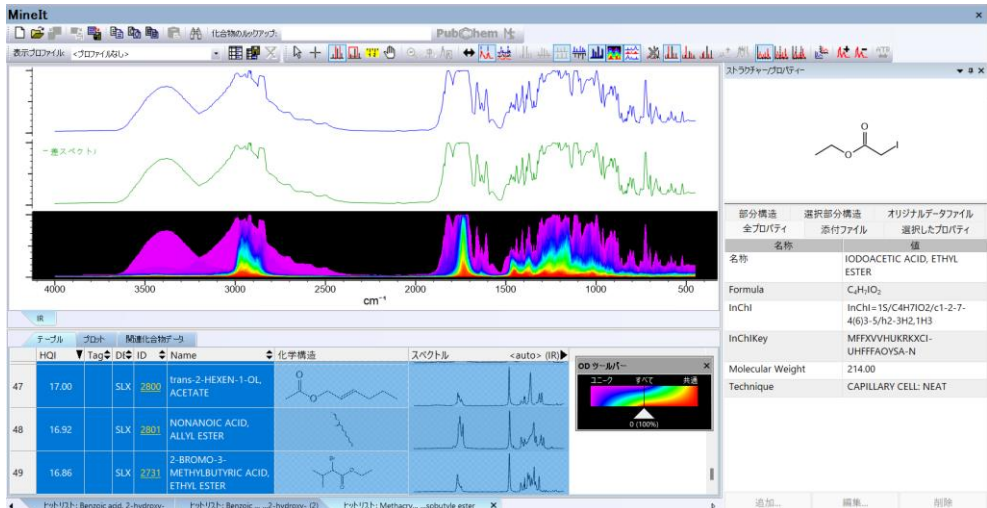
部分構造検索を行います

	アクション	結果
1	<p>SearchIt アプリケーションで Search Databases の下にある User-Select をクリックします。</p> <p>「Available for Searching」から、「IR - Sadtler Standards (Selected Subset) - Wiley (DB Code SLX)」を選択します。</p> <p>「Add」をクリックします。</p>	
2	<p>「Structure」をクリックします。</p> <p>「Open Spectrum or Structure」アイコンをクリックします。</p>	

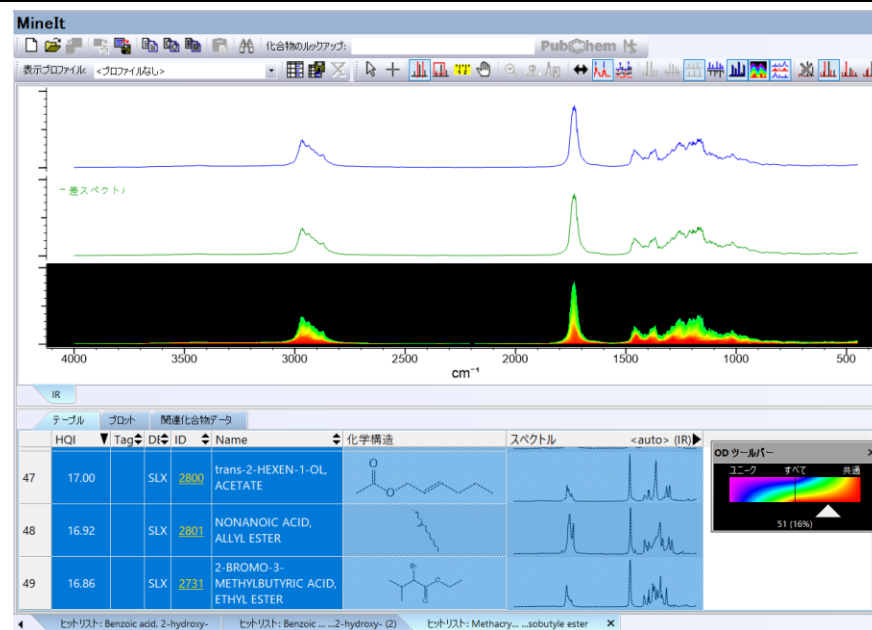
	アクション	結果
3	<p>C:\Users\Public\Documents\Wiley\KnowItAll\Samples\Structures フォルダに</p> <p>「Methacrylic acid, isobutyl ester」を選択します。</p>	構造が「Structure」タブに表示されます。
4	<p>「Search Mode」が「Similarity」に設定されており、デフォルトの「Tanimoto アルゴリズム」が使用されていることを確認してください。</p> <p>「Search」をクリックします。</p>	 <p>The screenshot shows the SearchIt application window. On the left, there are several search method options: 'Spectral FTIR', 'Spectral', 'Peak Information', 'Structure' (checked), and 'Proton/Name'. Under 'Structure', there are checkboxes for 'Stereochemistry', 'Relative Configuration', and 'Standardize Structure'. Below that is the 'Search Database' section with 'Custom Selection' (checked) and options for 'All Compounds', 'Pure Compounds', and 'Summary'. The 'Search Mode' section on the right shows 'Similarity' selected with the 'Tanimoto' algorithm. The 'Structure Edit' section has various options like 'Remove all elements', 'Remove all C and H', etc. The main search area on the right displays a chemical structure of methyl methacrylate isobutyl ester. At the bottom, there are search parameters like 'Number of hits: 50' and a 'Search' button.</p>

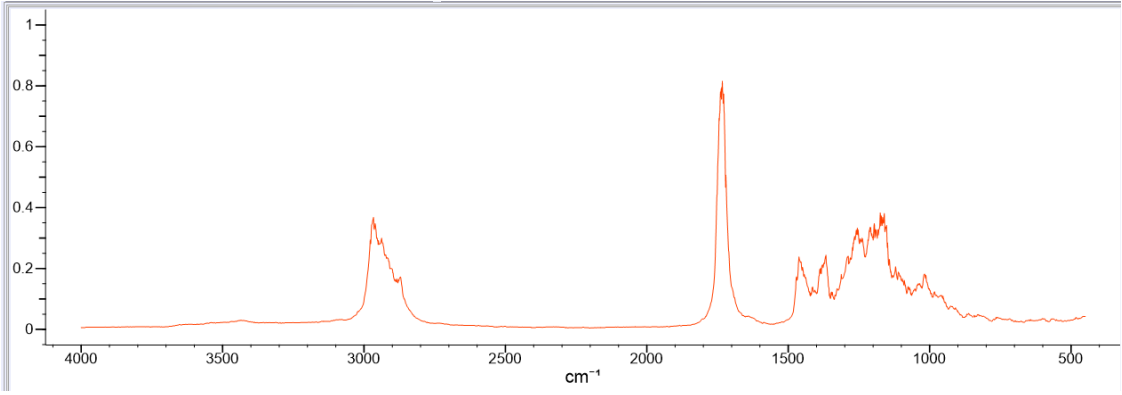
Minelt での結果を確認

	アクション	結果
1	<p>ヒットリスト内の 11 件のレコードをすべて選択します。</p> <p>注記: 最初のレコードを選択し、Shift キーを押しながら最後のレコードまでスクロールして選択します。</p>	<p>全てのスペクトルがスペクトルパネルに表示されます。</p> 
2	<p>OD ヒートマップツールバーのボタンをクリックします。</p>	

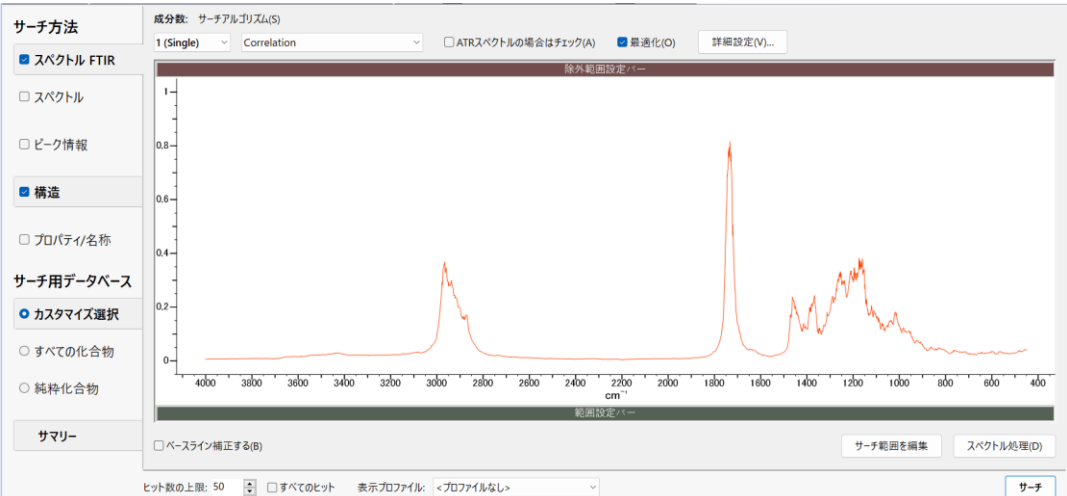
	アクション	結果																																						
3	<p>OD コンセンサスペクトルツール</p> <p>バーのボタン  をクリックします。</p>	<p>オーバーラップ密度ヒートマップの上にコンセンサスペクトルが表示されます。</p>  <p>The screenshot shows the Minelt software interface. At the top, there are two IR spectra (one blue, one green) overlaid on a density heatmap. Below the spectra is a table with columns for Retention Time (RT), Tag, ID, Name, Chemical Structure, and Spectrum. The table contains three entries:</p> <table border="1"> <thead> <tr> <th>RT</th> <th>Tag</th> <th>ID</th> <th>Name</th> <th>Chemical Structure</th> <th>Spectrum</th> </tr> </thead> <tbody> <tr> <td>17.00</td> <td>SLX</td> <td>2890</td> <td>trans-2-HEXEN-1-OL, ACETATE</td> <td><chem>CCCCC(=O)OCC</chem></td> <td></td> </tr> <tr> <td>16.92</td> <td>SLX</td> <td>2901</td> <td>NONANOIC ACID, ALLYL ESTER</td> <td><chem>CCCCC(=O)OCC</chem></td> <td></td> </tr> <tr> <td>16.86</td> <td>SLX</td> <td>2731</td> <td>2-BROMO-3-METHYLBUTYRIC ACID, ETHYL ESTER</td> <td><chem>CCCC(=O)OCC</chem></td> <td></td> </tr> </tbody> </table> <p>On the right side of the interface, there is a chemical structure diagram and a table of properties:</p> <table border="1"> <thead> <tr> <th>名称</th> <th>値</th> </tr> </thead> <tbody> <tr> <td>名称</td> <td>iodoacetic acid, ethyl ester</td> </tr> <tr> <td>Formula</td> <td>C₈H₁₆O₂</td> </tr> <tr> <td>InChI</td> <td>InChI=1S/C4H7IO2/c1-2-7-4(6)-5/h2-3H2,1H3</td> </tr> <tr> <td>InChIKey</td> <td>MFFXVHVURKXCI-UHFFFAQYSA-N</td> </tr> <tr> <td>Molecular Weight</td> <td>214.00</td> </tr> <tr> <td>Technique</td> <td>CAPILLARY CELL: NEAT</td> </tr> </tbody> </table>	RT	Tag	ID	Name	Chemical Structure	Spectrum	17.00	SLX	2890	trans-2-HEXEN-1-OL, ACETATE	<chem>CCCCC(=O)OCC</chem>		16.92	SLX	2901	NONANOIC ACID, ALLYL ESTER	<chem>CCCCC(=O)OCC</chem>		16.86	SLX	2731	2-BROMO-3-METHYLBUTYRIC ACID, ETHYL ESTER	<chem>CCCC(=O)OCC</chem>		名称	値	名称	iodoacetic acid, ethyl ester	Formula	C ₈ H ₁₆ O ₂	InChI	InChI=1S/C4H7IO2/c1-2-7-4(6)-5/h2-3H2,1H3	InChIKey	MFFXVHVURKXCI-UHFFFAQYSA-N	Molecular Weight	214.00	Technique	CAPILLARY CELL: NEAT
RT	Tag	ID	Name	Chemical Structure	Spectrum																																			
17.00	SLX	2890	trans-2-HEXEN-1-OL, ACETATE	<chem>CCCCC(=O)OCC</chem>																																				
16.92	SLX	2901	NONANOIC ACID, ALLYL ESTER	<chem>CCCCC(=O)OCC</chem>																																				
16.86	SLX	2731	2-BROMO-3-METHYLBUTYRIC ACID, ETHYL ESTER	<chem>CCCC(=O)OCC</chem>																																				
名称	値																																							
名称	iodoacetic acid, ethyl ester																																							
Formula	C ₈ H ₁₆ O ₂																																							
InChI	InChI=1S/C4H7IO2/c1-2-7-4(6)-5/h2-3H2,1H3																																							
InChIKey	MFFXVHVURKXCI-UHFFFAQYSA-N																																							
Molecular Weight	214.00																																							
Technique	CAPILLARY CELL: NEAT																																							

- 4 OD ツールバーを約 50 (16%) に移動
します。



	アクション	結果
5	再度 OD ヒートマップツールバーのボタンをクリックして、表示をオフにします。	<p>コンセンサスペクトルは表示されたままです。</p> 

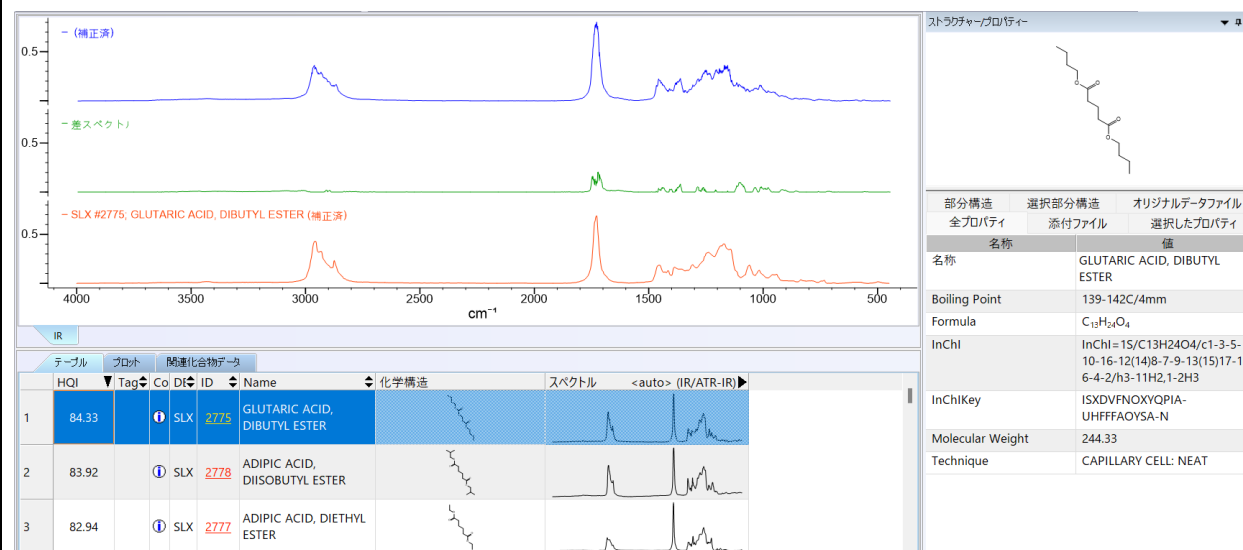
コンセンサスペクトルを使用して検索を行います

	アクション	結果
1	Transfer to バーで SearchIt をクリックしてください。	<p>コンセンサスペクトルは IR スペクトル検索タブに表示されています。</p>  <p>検索方法: サーチアルゴリズム(S) 1 (Single) Correlation <input type="checkbox"/> ATRスペクトルの場合はチェック(A) <input checked="" type="checkbox"/> 乗算化(O) 詳細設定(V)...</p> <p><input checked="" type="checkbox"/> スペクトル FTIR <input type="checkbox"/> スペクトル <input type="checkbox"/> ピーク情報</p> <p><input checked="" type="checkbox"/> 構造 <input type="checkbox"/> プロパティ/名称</p> <p>サーチ用データベース <input checked="" type="radio"/> カスタマイズ選択 <input type="radio"/> すべての化合物 <input type="radio"/> 純粋化合物</p> <p>サマリー <input type="checkbox"/> ベースライン補正する(B)</p> <p>ヒット数の上限: 50 <input type="checkbox"/> すべてのヒット 表示プロファイル: <プロフィールなし> <input type="button" value="サーチ"/></p>

2 「Search」 ボタンをクリックしてください。

注記：IR - Sadtler Standards (Selected Subset) - Wiley データベースはスペクトルパネルに表示されたままです。

Minelt アプリケーションの新しいタブで結果が表示されます。



Methacrylic acid, isobutyl ester と構造的に類似した化合物のヒットリストが表示されました。