天然有机质分子分析平台（nmc）V2.0

说明书

目　录

1. 软件概览　1

2. 软件开发环境　2

3. 主要界面介绍　2

3.1 主功能列表界面　2
3.2 分子解析界面　3
3.3 分子信息界面　6
3.4 分子结构界面　21
3.5 活性信息界面　23
3.6 工作路径和案例数据　25

4. 操作方法　26

4.1 软件安装　26
4.1.1 EXE 版本　26
4.1.2 PCODE 版本　26
4.1.3 其他注意事项　27
4.2 有机质的分子解析运行　27
4.3 有机质的分子信息计算　28
4.3.1 案例 1：导入未处理样本并显示范氏图（3D）　29
4.3.2 案例 2：导入结果数据并显示样本 S2 独特分子 KMD（CH_2）图　30
4.4 有机质的分子网络分析　30
4.5 有机质的分子转化分析　31
4.6 有机质的分子结构检索　32
4.6.1 案例 1：选择单个分子　33
4.6.2 案例 2：选择全部分子　33
4.7 有机质的分子活性计算　33

1. 软件概览

nmc 平台是一款可实现从天然有机质高分辨率质谱数据解析分子、提取信息、查询结构和估算活性的软件平台。nmc 平台可实现四项功能：（1）从高分辨率质谱（包括傅里叶变化离子回旋共振质谱仪 FT-ICR-MS、轨迹陷阱质谱仪 Orbitrap-MS、飞行时间质谱仪 TOF-MS 所输出的数据）数据中**解析**出每个质荷比（m/z）所对应的有机分子式；（2）基于有机分子式计算该分子的元素组成特征、可能分子种类、不饱和指数、同系物、分子网络特征和分子转化潜力等**信息**；（3）通过 PubChem 数据库检索分子式所对应分子**结构**的名称、符号和图示等多种呈现形式；（4）利用结构活性预测软件（EPI Suite）和分子结构符号表达式（SMILES）对分子的**活性**参数进行估算和统计分析。

本软件的理论和计算涉及通过 m/z 值—有机分子数据库解析有机质分子式、分子式元素和种类组成信息、结构查询和活性估算等有机质分子分析的相关内容，由于有机质分子数量庞大，利用 Excel 等传统数据统计软件需要耗费大量时间，而大多数研究者并不掌握 MATLAB 语言，也不熟悉分子式结构的查找，更无法通过分子结构获得其潜在的活性数据。可见，从 m/z 值到其潜在活性特征需要的工作量十分巨大，需要一款集成软件进行规范的分析。本平台以 MATLAB GUI 开发操作界面，调用了 PubChem 数据库和 EPI Suite 软件，并集成上述数据进行了深度分析。应用范围可囊括所有高分辨率质谱数据，实现天然有机质分子结构与活性分析。

图 1 nmc 的主要功能划分

nmc 的主要功能模块如图 1，包括主功能列表界面与分子解析、分子信息、分子结构和活性信息 4 个子界面。其中，后两个子界面分别需要电脑联网且能够访问 PubChem 网站以及安装有美国环保局（USEPA）开发的 EPI Suite 软件（链接：https://www.epa.gov/tsca-screening-tools/download-epi-suitetm-estimation-program-interface-v411），**针对 EPI Suite 软件，需要安装 2017 更新包**（EPI Suite updated files 2017.zip (zip)），按照上述网页介绍进行安装。

2. 软件开发环境

本软件利用 Matlab 2018b GUI 进行了用户界面的开发，代码为 MATLAB 语言。本软件仅含 64 位版本，软件包含有案例数据供初学者学习。

3. 主要界面介绍

3.1 主功能列表界面

3.1.1. 主功能列表界面（图 2）罗列出本软件的四个功能模块，可分别导引至“分子解析界面”、“分子信息界面”、“分子结构界面”和“活性信息界面”。通过“×”按钮可关闭“主功能列表界面”。“主要功能说明”介绍了本软件的一些基本信息、特色功能以及著作权保护等内容。

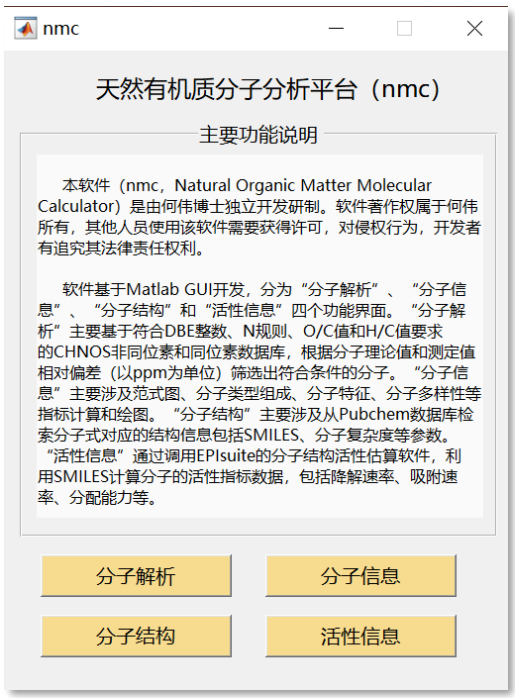


图 2 主要功能列表界面

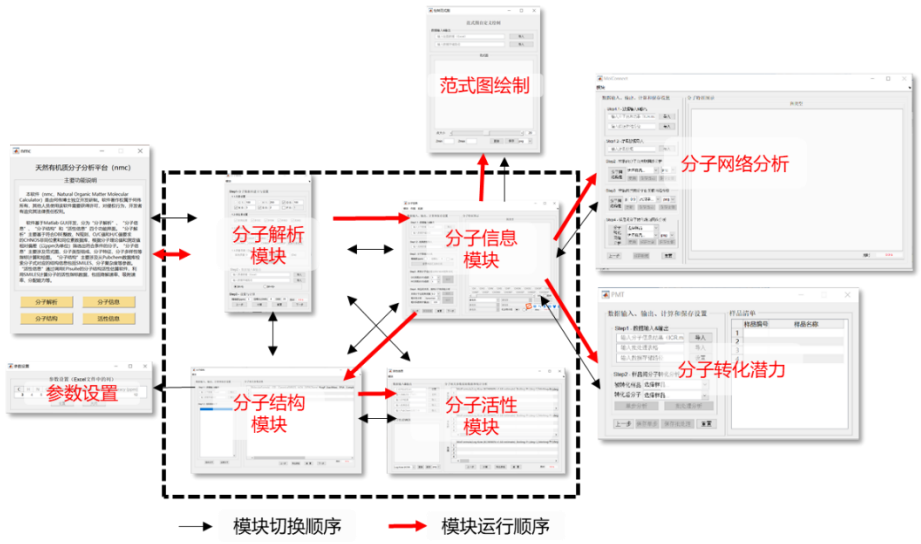


图 3 nmc 各功能界面的关联

3.1.2. 主功能列表界面通过图 3 所示的各功能界面的关联图实现与其他界面的切换，模块运行顺序一般按照红色箭头进行，即先进行分子解析，后面可分别进行分子信息计算（含分子网络分析和转化分析）和分子结构检索，在分子结构基础上，可进行分子活性信息的估算。

3.2 分子解析界面

3.2.1. nmc 分子解析界面分为“分子数据库建立与设置”、“数据输入&输出”和“设置与计算”三个部分。“分子数据库建立与设置”根据 C、H、O、N、S 和 P 的最大数量建立局部分子数据库，考虑到有机分子至少包含 C 和 H，这两个元素是必选项，其他元素为可选项。对于 N、S 和 P 等杂原子，其默认数量设置分别为 3、2 和 1，考虑到 CHONSP 可以构建 16 种组合元素的分子，但有些分子类型并非研究所关注的，**为了加速分子库构建以及后续的数据分析，这里可以有选择性的筛选分子的元素组成情况、以及质量范围情况**。“数据输入&输出”主要是实现对导入数据和输出路径进行选择，“设置与计算”通过设置精确度（数据库理论分子量与测定分子量之间的相对偏差，以百万分之一（ppm）为单位）、信噪比（SNR）和等双键当量（DBE）等参数来筛选出符合预设条件的分子来构建分子库。

3.2.2. 本软件输出的含 CHONSP 元素（含稳定同位素，^{13}C, ^{18}O, ^{15}N 和 ^{34}S）有机分子数据库 CHNOS1000（图 5），该数据库有机分子分子量范围可自行设置如 0-1000 Da，该分子库包括不含同位素有机分子子数据库（basic）、含 ^{13}C 同位素值有机分子子数据库（C13）、含 ^{18}O 同位素值有机分子子数据库（O18）、含 ^{15}N 同位素值有机分子子数据库(N15)和含 ^{34}S 同位素值有机分子子数据库(S34)。考虑到上述稳定同位素丰度相对于主量同位素均较低，同位素有机分子数据库中均仅含有一个同位素分子，而同时在一个分子中出现两个同位素的概率极低，因此本软件数据库中未考虑包含 2 个以上稳定同位素的有机分子。数据库数据集的

各列分别为理论分子量、等双键当量（DBE）和六种元素个数。

图 4 nmc 分子解析界面

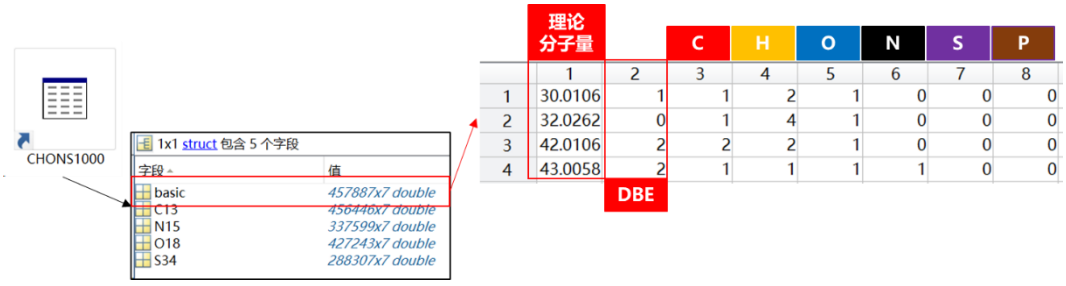


图 5 有机分子数据库构成及数据集解释

3.2.3. 有机分子数据库的建立遵循如下规则：

- 各元素数量范围遵循 $C_{1\text{-}c}H_{1\text{-}h}O_{0\text{-}o}N_{0\text{-}n}S_{0\text{-}s}P_{0\text{-}p}$，默认值分别为 100、200、100、3、2 和 1。
- 分子量范围可控制，默认 0-1000 Da

- DBE（C + 0.5×N + 0.5×P + 1 –0.5×H – 0.5）为非负整数
- N/C≤1
- H/C≥0.3
- O/C≤1
- 数据库计算的是中性分子
- 剔除掉非目标元素组成的有机分子类型

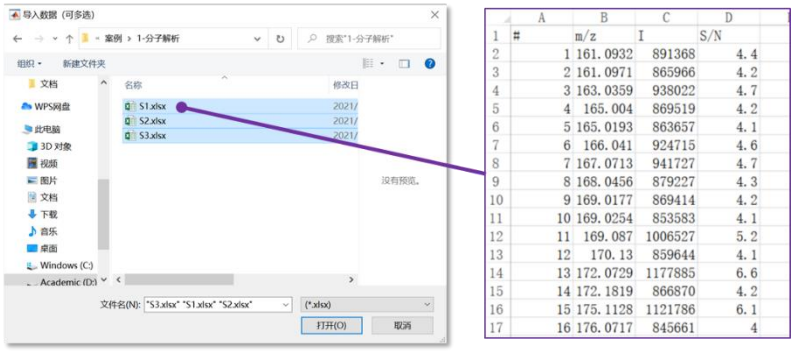
3.2.4. 分子比对、筛选和解析遵循如下规则：

- 针对带电荷分子，如 ESI 负模式（以形成负电荷离子为主）的$[M-H]^-$的负离子，在反算中性分子时需要加上氢原子质量（1.00782503207）并减去电子质量（0.000548597）；又如 ESI 正模式（以形成正电荷离子为主）的$[M+H]^+$正离子，在反算中性分子时需要减去氢原子质量并加上电子质量。
- 根据 DBE 设置的最大值（在一定程度上可以排除高聚合度分子），剔除分子数据库中大于该设定值的分子。
- 将某有机质分子所有 m/z 数值与非同位素子数据库比对，计算精确度；确定满足精度要求的待选有机分子；在待选分子中遵循杂原子数量最少、精度值最小的原则确定有机分子的最可能的分子式。另外，再次剔除无法满足条件的分子。
- 按照上述算法将被剔除的 m/z 数据与同位素子数据库比对，根据设定精确度确定含有同位素的有机分子，且该分子在之前确定的非稳定同位素有机分子族中存在；当同一个 m/z 对应 2 个满足条件的同位素有机分子时，按照 CONS 的顺序选取最前面的一个。**分子解析功能所利用分子数据库必须包括非同位素、13C、15N、18O 和 34S 子库。**
- 解析出来的分子的元素组成类型小于或等于创建分子数据库包含的类型数量。

3.2.5. 数据导入和路径选择：高分辨率质谱仪一般会产生带有质荷比（m/z）、强度（I）和信噪比（S/N）的数据，分子解析模块需要如图 6 所示的形式组织数据，数据可存入 Excel 文档（xlsx 或 xls）和逗号分隔值文档（csv）中；本模块可以多选数据文档进行批量分子解析操作。导入数据输出路径的默认数据存储文件夹名称是“结果”，该文件夹为软件安装时可见的一个文件夹，建议不要删除。

3.2.6. 设置与计算：可设置解析分子的精确度（ppm），默认值为 1，即解析分子的理论值与实测值之间的相对偏差绝对值小于或等于 1，信噪比（SNR）和 DBE，默认值分别为 4 和 20；通过“计算”来运行分子解析模块，在运行过程中，会依次出现如图 7 所示的不含同位素有机分子和含同位素有机分子的解析进

度条；通过“上一步”回到主功能列表界面；通过“重置”复位界面上的所有数据输入信息和结果输出信息；通过“下一步”进入“分子信息”界面。



	A	B	C	D
1	#	m/z	I	S/N
2	1	161.0932	891368	4.4
3	2	161.0971	865966	4.2
4	3	163.0359	938022	4.7
5	4	165.004	869519	4.2
6	5	165.0193	863657	4.1
7	6	166.041	924715	4.6
8	7	167.0713	941727	4.7
9	8	168.0456	879227	4.3
10	9	169.0177	869414	4.2
11	10	169.0254	853583	4.1
12	11	169.087	1006527	5.2
13	12	170.13	859644	4.1
14	13	172.0729	1177885	6.6
15	14	172.1819	866870	4.2
16	15	175.1128	1121786	6.1
17	16	176.0717	845661	4

图 6 样品数据导入

3.2.7. 本界面导出的结果数据文件见图 8，“分子解析结果+时间”文件夹包含了所有输入数据对应的输出结果文件，在结果文件中，主要涉及分子解析相关的数据包括元素组成（Category）、同位素类型（Isotope，空白表示不含同位素、其他同位素用 13C、15N、18O 和 34S 进行了标注）、元素个数（C、H、N、O、S、P）、分子量理论值（Theo）、分子量实测值（m/z）、信号强度（Intensity）、信噪比、精确度（Acc）、分子类别（Class）、双键当量（DBE）、分子式（Formula）等。

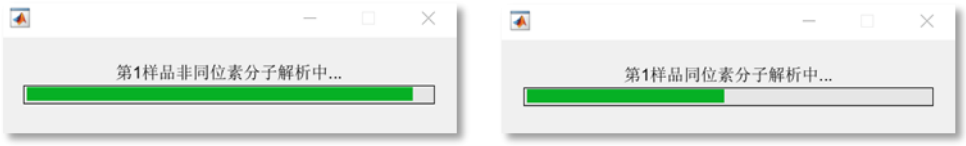
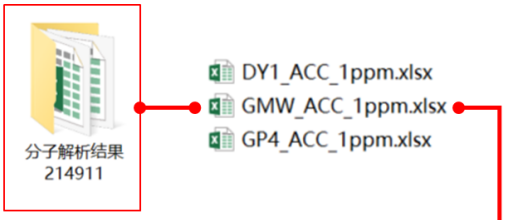


图 7 点击分子解析界面“计算”后不含同位素有机分子和含同位素有机分子的解析进度条



Category：元素组成
Isotope：同位素类型
CHNOS：元素个数
Theo：分子量理论值
m/z：分子量实测值
Intensity：信号强度

s/n：信噪比
Acc：精确度（ppm）
Class：分子类别
DBE：双键当量
Formula：分子式

Category	Isotope	C	H	N	O	S	Theo	m/z	Intensity	s/n	Acc	Class	DBE	Formula
CHONS		8	16	1	2	1	190.0907	190.0909	4378138	28.3	-0.88193	O2N1S1	1	C8H16O2N
CHON		7	15	2	4		191.1037	191.1037	80919048	557	0.186108	O4N2	1	C7H15O4N
CHONS		7	14	1	1	2	192.0522	192.0522	2839941	17.6	0.300274	O1N1S2	1	C7H14O1N
CHO		9	11		5		199.0612	199.0613	2964196	18.2	-0.40667	O5	4	C9H11O5

图 8 分子解析界面导出结果

3.3 分子信息界面

3.3.1. 分子信息界面包含“数据输入、输出、计算和保存设置”(3.3.1-3.3.7)以及“分子特征图示”(3.3.8-3.3.12)两个部分。其中，前者涉及四个步骤（图 9）：（1）数据输入和输出设置、以及结果数据的导入，（2）分子筛选与分类，（3）

针对多样品情况对所有分子信息进行统计、以及将范氏图以分子数量和累积强度为 Z 轴进行三维呈现，（4）样品间共同分子和独特分子分析、以及分子间和样品间的相关性分析和聚类分析。后者主要通过多级关联弹出菜单方式选择要显示或者输出有机质分子信息的图示。分子信息界面的两个拓展功能是“分子网络”和“分子转化”。

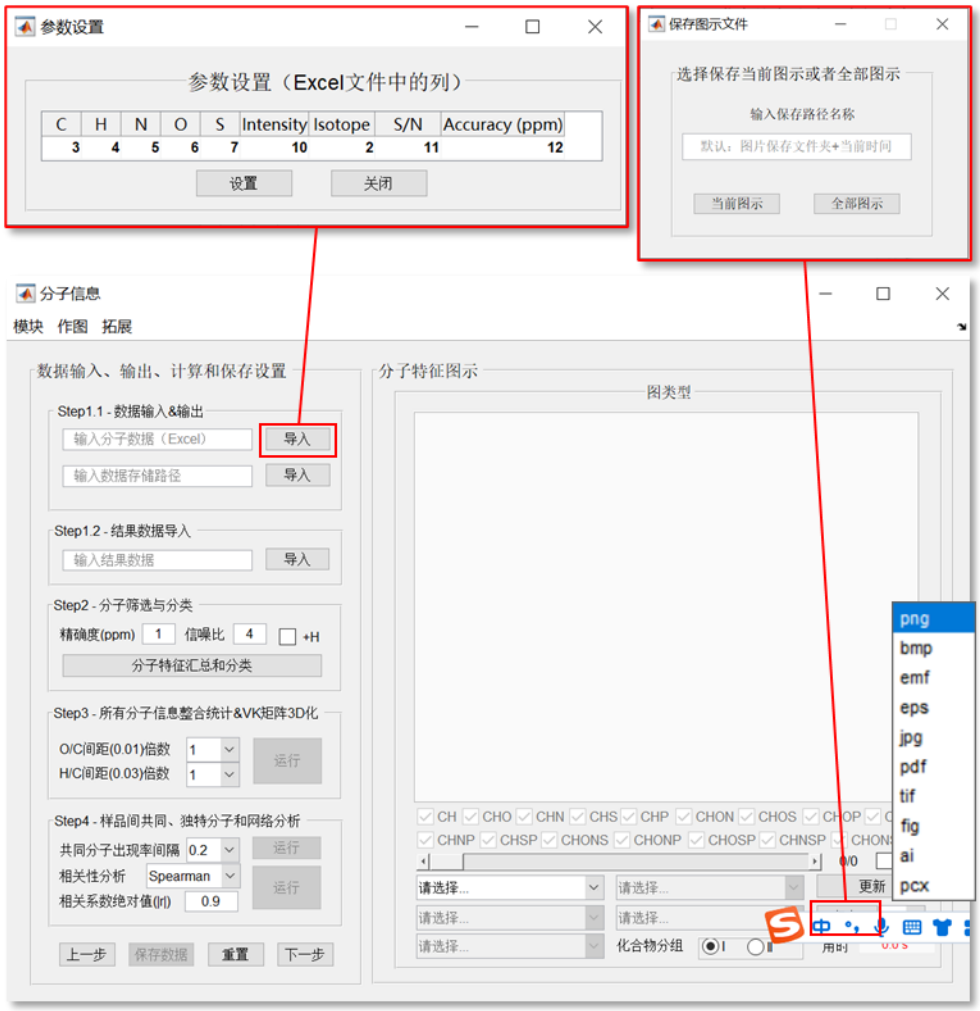


图 9 分子信息界面

3.3.2. 数据输入&输出：分子信息模块需要如图 8 分子解析模块所产生的数据文档形式组织数据，或者直接使用该模块产生的数据，数据可存入 Excel 文档（xlsx 或 xls）和逗号分隔值文档（csv）中；本模块可以多选数据文档进行分子信息的批量提取；由于后续功能会涉及到样品间的分析，建议不要只输入一个样品，如果只有一个样品，可以考虑复制获得一个副本，一同进行后续数据处理。点击“导入”数据后，会弹出如图 9 所示的参数设置界面，对输入数据文档中 CHNOSP 元素、信号强度等进行所在列号进行设置，如 C 数量位于 Excel 的 A 列，则这里输入 1。数据输出导入路径参考 3.2.5。

3.3.3. 结果数据导入：本软件可以直接读取通过 3.3.7 步骤所保存数据（mat 格式），命名为 ICR+(acc*sn*)+时间。导入结果数据后，无需做后续的分析，即可通过弹出菜单进行图示的显示和保存。如果分析参数有所调整，也可以按照

3.3.4 至 3.3.7 的步骤做参数调整后的数据分析。如果结果数据较大，读取数据会耗费一定的时间，待弹出如下对话框时，则表明数据导入成功，该对话框 1 秒后自动关闭，不要手动关闭。

3.3.4. 分子筛选与分类：由于本界面的输入数据既可以来自分子解析界面产生的结果数据，也可来自其他软件输出数据，因此，需要对数据进行进一步的筛选，本次筛选开放的约束条件有精确度（见 3.2.1 定义）和信噪比，默认情况下，精确度数值为 1 ppm，信噪比为 4。此外，一般质谱仪器分析采用 ESI 负模式，有机分子的 m/z 数据是轰击掉 H 后带负电荷的离子，因此，要还原到轰击前的中性分子，就需要通过点选 H 复选框（**注意：由于 nmc 分子解析模块输出的是中性分子，分子信息模块 H 复选框默认为未勾选状态**）。未开放的筛选规则除分子解析模块 3.2.3 所涉及的外，还包括 S/C 值小于或等于 0.2。分子信息包括 H/C、O/C、N/C、S/C、DBE、DBE-O（扣除氧元素数量的双键当量）、AI（芳香性指数）、AI_{mod}（改进的芳香性指数）、AI_{con}（评估羧基含量水平的芳香性指数）、NOSC（碳的名义氧化态）、X_c（芳香度）、ExactMass（理论分子量）、KMD（Kendrick 质量亏损，包含 CH_2、CO_2、H_2、H_2O、O、SO_2、NH_2 和 C_6H_5），各指标计算方法如下：

- H/C、O/C、N/C、S/C、P/C 元素数量比值
- DBE，见 3.2.3
- DBE-O，DBE-O = DBE – O
- AI，AI = (1 + C – O – S – 0.5×(H +N)) / (C – O – S – N – P)
- AI_{mod}，AI_{mod} = (1 + C – 0.5×O – S – 0.5×(H +N)) / (C – 0.5×O – S – N – P)
- AI_{con}，AI_{con} = (1 + C – $(COOH)_n$ – S – 0.5×H – 0.5×N) / (C –$(COOH)_n$ – S – N – P)
- NOSC，NOSC = 4 – (4×C + H – 3×N – 2×S – 2×O) / C
- X_c，X_c = (3 × (DBE – O – S) – 2) / (DBE – O – S)
- KMD_h，KMD_h = Round(m/z × NomMass / ExactMass) – m/z × NomMass / ExactMass，NomMass 是 CH_2、CO_2、H_2、H_2O、O、SO_2、NH_2、C6H5 等同系物结构的名义质量，分别为 14、33、2、18、16、64、16、77 Da，ExactMass 是它们的理论质量，分别为 14.01565、43.989830、2.015650、18.010565、15.994915、63.961902、16.018724084、77.039125185 Da，Round()是取整函数，m/z 是有机分子理论质量；软件会对所有分子计算

KMD，但只有分子中含上述同系物结构，该数值才是有意义的。

分子分类有两个方法，一是按照元素组成分类，包含 CHO、CHN、CHON、CHOS、CHONS、CHONP 等多个类型；二是依据 O/C、H/C、DBE、AI_{mod} 和 N 按照分子特征分类，包括两个常见分类，如下：

- 分类 I
  - I1：稠芳香烃物质（Condensed aromatic structures，CAS），$AI_{mod} > 0.67$
  - I2：芳香烃物质（Aromatic structures，AS），$0.5 \leq AI_{mod} \leq 0.67$
  - I3：低氧高不饱和物质（Highly unsaturated, low oxygen，HUSLO），$AI_{mod} < 0.5$，$H/C < 1.5$，$O/C < 0.5$
  - I4：高氧高不饱和物质（Highly unsaturated, high oxygen，HUSHO），$AI_{mod} < 0.5$，$H/C < 1.5$，$0.5 \leq O/C \leq 0.9$
  - I5：脂肪类物质（Aliphatics），$1.5 \leq H/C \leq 2$ ，$O/C < 0.9$，$N = 0$
  - I6：多肽类物质（Peptides），$1.5 \leq H/C \leq 2$ ，$O/C < 0.9$，$N > 0$
  - I7：糖类物质（Sugars），$O/C > 0.9$
  - I8：其他分子（Others），除去上述类别以外的其他分子
- 分类 II
  - II1：类脂肪物质（Lipid-like），$1.5 \leq H/C \leq 2$，$O/C < 0.3$
  - II2：类蛋白质物质（Protein-like），$1.5 \leq H/C \leq 2.2$，$0.3 \leq O/C \leq 0.67$，$N > 0$
  - II3：类木质素物质（Lignin-like），$0.7 \leq H/C < 1.5$，$0.1 \leq O/C \leq 0.67$
  - II4：类碳水化合物质（Carbohydrate-like），$1.5 \leq H/C \leq 2.4$，$O/C > 0.67$
  - II5：不饱和烃类物质（Unsaturated hydrocarbon），$0.7 \leq H/C < 1.5$，$O/C < 0.1$
  - II6：稠芳香烃类物质（Condensed aromatic ring structure），$0.2 \leq H/C < 0.7$ ，$O/C \leq 0.67$
  - II7：单宁类物质（Tannins-like），$0.6 \leq H/C < 1.5$ ，$O/C > 0.67$
  - II8：其他分子（Others），除去上述类别以外的其他分子
- 富含羧基的脂环族物质（Carboxylic-rich alicyclic molecules，CRAM），$0.3 \leq DBE/C \leq 0.68$，$0.77 \leq DBE/O \leq 1.75$，$0.2 \leq DBE/H \leq 0.95$

根据上述分类规则分类，并计算各分子组中有机分子的数量、数量百分比、以及分子强度加权百分比。

该功能运行过程中会显示如图 10 所示的进度条，完成后，会出现完成提示对话框。

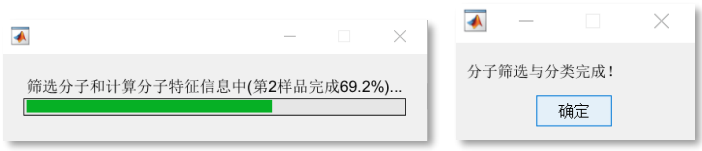


图 10 分子筛选与分类功能进度条与任务完成对话框

3.3.5. 所有分子信息整合统计&VK 矩阵 3D 化：本功能主要是对上述数据进行整合和统计分析，包括含奇数个氢有机分子数量与含偶数个氢有机分子数量的比值、3.3.4 分子信息指标的信号强度加权均值、基于数量和相对信号强度的范氏图三维矩阵化、分子多样性分析、以及对多样品的分子信息指标和不同类型分子的统计分析（含均值、标准偏差、最小值、最大值、中位数、第 25 百分位数和第 75 百分位数）。在本功能中，用户可以设置的参数是范氏图 3D 矩阵化时 O/C 和 H/C 的间距，即在制作 3D 矩阵时，需要对落入固定范围 O/C 和 H/C 的有机分子数量或者相对强度进行累加，如果固定范围越大，进入该范围的分子数量就越大，当然固定范围越大，矩阵的分辨率就约小。本功能中 O/C 和 H/C 的最小间距为 0.01 和 0.03，可设置其他间距分别为最小间距的 2、5、10 和 20 倍。默认采用的是最小间距。范氏图 3D 矩阵化的数据可以直接用于 efc 软件的 PARAFAC 分析模块进行进一步的分析，提取不同样品的共同特征并获得特征得分。分子多样性指标包括样品分子数量、改进的 Chao1 多样性指数（Chao1M）和 shannon 多样性指数。

- 改进的 Chao1 多样性指数：F1 为仅出现在本样品中的有机分子数量，F2 为出现在包括本样品在内的 2 个样品的有机分子数量，FF 为样品中所有分子的数量，则 Chao1M = FF + F1×(F1-1) ×0.5/(F2+1)
- shannon 多样性指数：piv 为某分子的相对信号强度，则 shannon = Σpiv×ln(piv)

在本功能执行过程中会出现的运算提示对话框如图 11 所示。

图 11 所有分子信息整合统计&VK 矩阵 3D 化功能进度条与任务完成对话框

3.3.6. 样品间共同和独特分子分析：本功能只能在输入数据样品数量超过一个才能使用。样品间共同分子主要包括在多个样品里出现的共同分子和两两分子间的共同分子。后者是前者的特殊情况，即在两个样品里出现的共同分子。为了便于后续说明，这里定义 N 为样品数量，则出现在 n 个样品中的分子为共同分子，对应的检出率（DF）计为 n/N。当 DF=1 时，即，在所有样品中均出现的共同分子。一般情况下，检出率越高，共同分子数量越少，本功能开放了对 DF 间距的设定，包括 0.2、0.1、0.05 和 0.01，若设定为 0.2，则 DF 为 0.2、0.4、0.6、0.8 和 1 的共同分子将会被提取，建议根据样品数量来设定 DF 间距，一般以 1/N 来定义间距，若有 10 个样品，则间距可定义为 0.1。点击第一个“运行”按钮后，将会对设定 DF 间距下不同检出率的共同分子进行归类分析；同时，还会统计仅在每个样品内出现的独特分子；此外，本功能还会计算两个样品之间所有分子、共同分子、各自独特分子，并能进行归类分析。在本功能执行过程中会出现的运算提示对话框如图 12 所示。

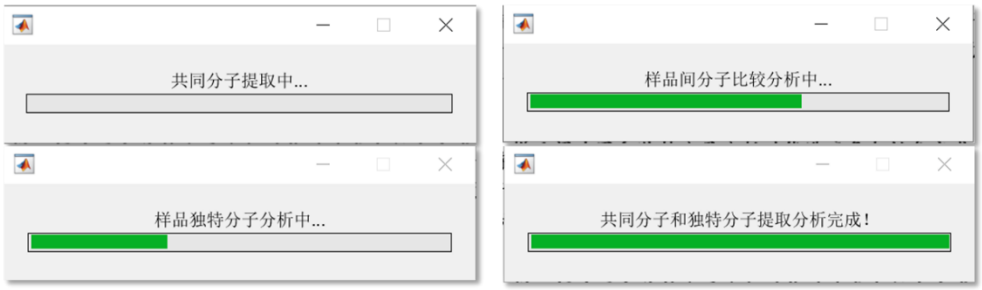


图 12 样品间共同和独特分子分析功能进度条与任务完成对话框

第二个“运行”按钮的功能是进行分子网络分析数据的整理，所整理数据符合 Gephi 软件要求，为了构建各分子之间的联系，采用了基于最大 DF 值时各样品共同分子信号相对强度数据进行相关性分析，相关性分析方法包括 Pearson、Spearman 和 Kendall，相关系数绝对值作为分子间连接性（网络分析）的阈值。

3.3.7. 数据保存：点击“保存数据”后，本软件会输出所有关键结果数据，如图 13 所示。

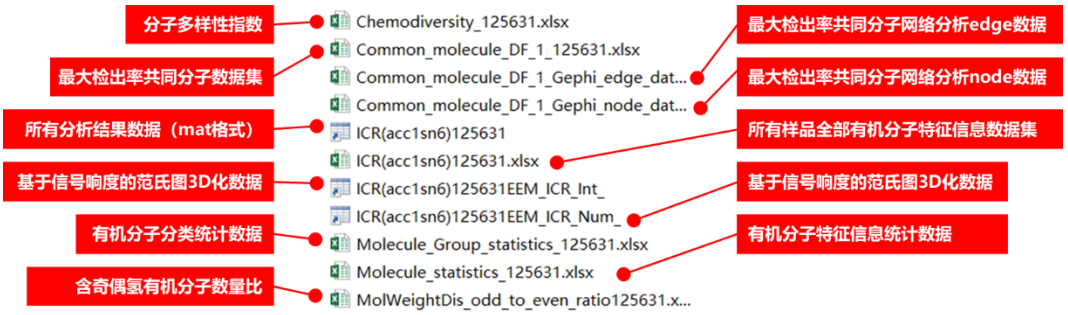


图 13 主要输出结果数据

3.3.8. 分子特征图示：该部分除重要多级关联菜单外，还包括不同元素组成分子选择和图例选择，一般可以通过勾选某类分子而只显示该类分子的范氏图。

点击“更新”时，“图类型”会生成当前的多级关联菜单当前的选择项，同时绘制相应的图示，点击“保存”时，会出现如图 8 所示的保存图片提示对话框，点击“当前图示”，则会保存当前图示，点击“全部图示”或者直接点“×”退出该界面，都会保存最后一级菜单中的所有选项对应的图示。保存的图示类型包括矢量图（pdf、fig、eps 和 svg）和位图（png、jpeg、tiff 和 bmp）。在保存图示的同时也会保存制作该图所需的数据，以便于用户制作符合自己习惯和风格的图示。保存图示及其绘图数据的默认文件夹命名为“图片保存文件夹+时间”。

3.3.9. 分子筛选与分类分析：第二级菜单选项包括质谱图、范氏图、分子特征、分子类型、氮硫分子组成、特征关系等。

- 质谱图：以分子量为横坐标，分别以相对信号强度（按照 m/z 最大信号值进行归一化）、分子数量（标称 m/z 质量下的分子个数）、元素组成分类（见 3.3.4 介绍）中各类分子数量、分子特征分类（I 和 II 类，见 3.3.4 介绍）中各类分子数量为纵坐标的四个图示（图 14），分子特征两个分类方式通过“化合物分组”进行控制。

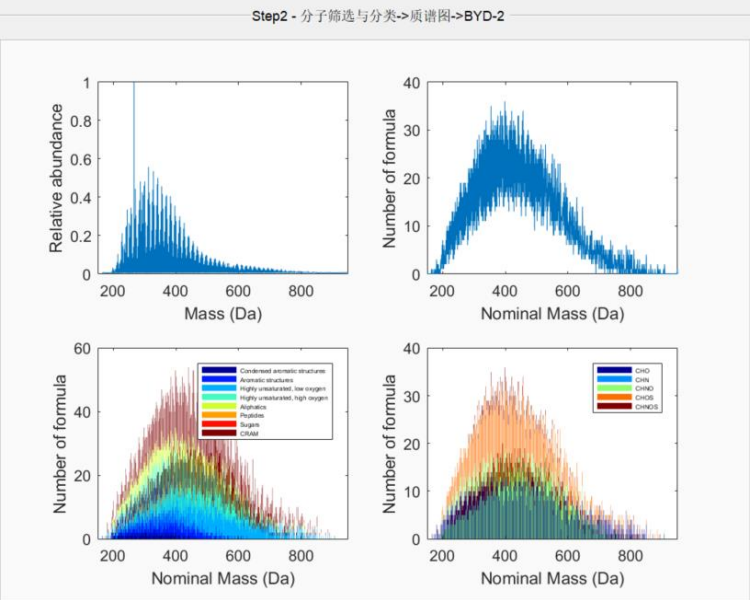


图 14 质谱图

- 范氏图：第三级菜单中可选不同的特征信息指标（见 3.3.4 节）为 Z 轴数据，以点的颜色梯度变化指示数值大小（图 15）。除了范氏图主图外，其他三个辅图分别是沿着 O/C 轴统计的不同元素组成有机分子的数量叠加柱状图、沿着 H/C 轴统计的不同元素组成有机分子的数量叠加柱状、以及不同元素组成有机分子的选定指标的箱式图。

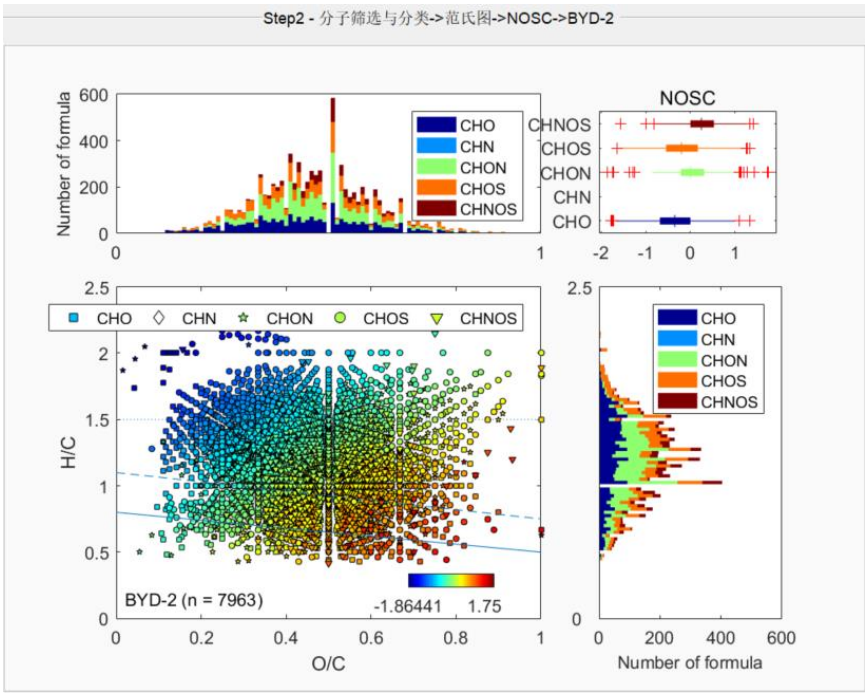


图 15 范氏图

- 分子特征：以某个特征分子信息为统计指标，如 NOSC，按照元素组成分类和分子特征（I 和 II 类全称见 3.3.4 节）分类分别统计不同类型有机分子族某统计指标的概率分布图和箱式图（图 16）。

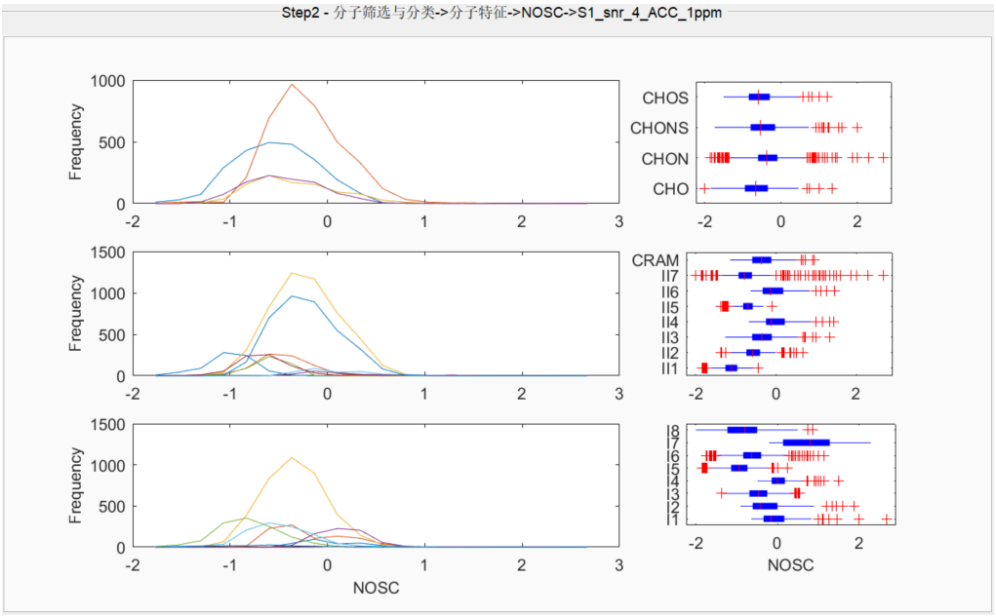


图 16 分子特征概率分布和箱式图

- 分子类型：按照元素组成分类和分子特征分类（I 和 II）分别统计不同类型有机分子的数量、数量占比和累积信号强度占比（图 17）。

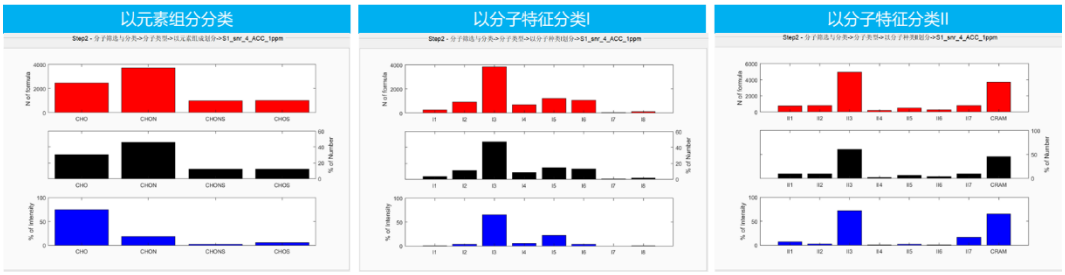


图 17 分子类型图

- 氮硫分子组成：依据 DBE 和氮硫元素，对含氧有机分子进行分类统计，可得到随着氧元素数量增加情况下，不同 DBE 的含硫或氮的含氧有机分子数量的叠加柱状图（图 18）。

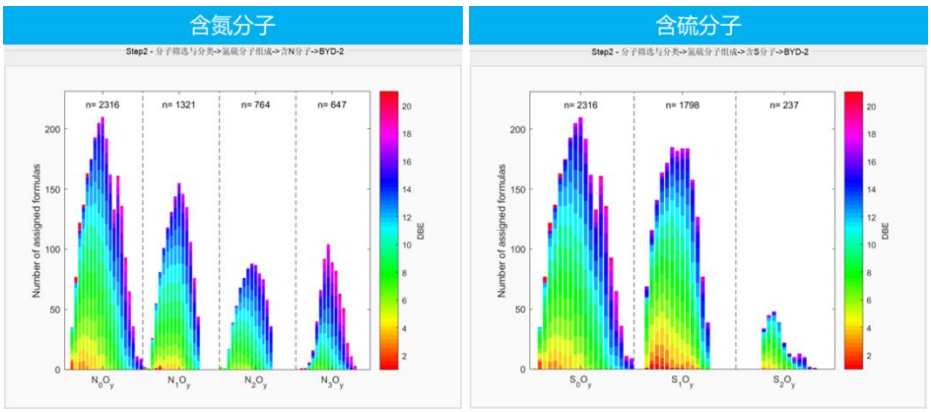


图 18 含氮硫分子图

- 特征关系：对 3.3.4 所列的有机质分子特征信息指标之间的关系进行展示，如图 19 所示为有机分子理论质量与 DBE-O 的关系。

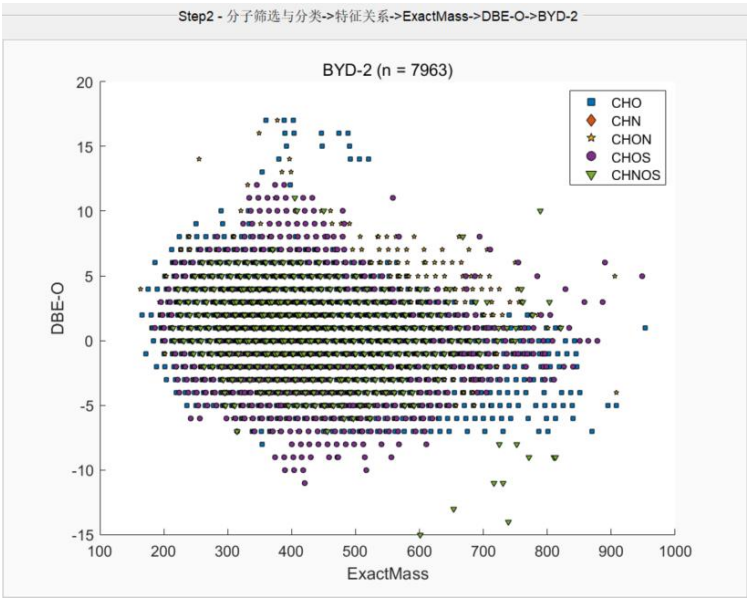


图 19 特征信息指标之间关系图

3.3.10. 所有分子信息整合统计分析：第二级菜单包括范氏图（3D）和分子多样性指数。

- 范氏图（3D）：将 3.3.5 所计算的基于分子数量和相对强度的三维化范氏图以等值线图的形式进行显示（图 20）。

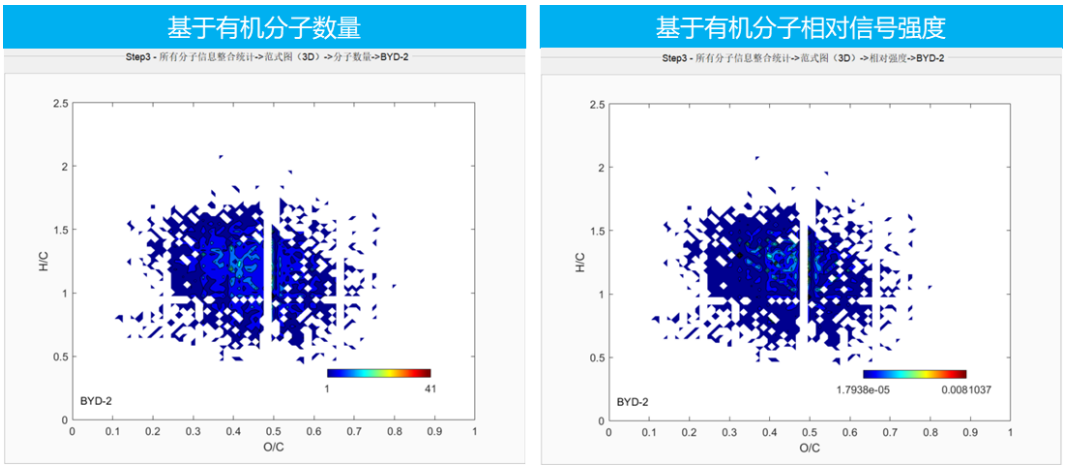


图 20 3D 形式范氏图

- 分子多样性指数：将 3.3.5 所计算的多样性指数（Chao1 和 shannon）以柱状图的形式显示（图 21）。

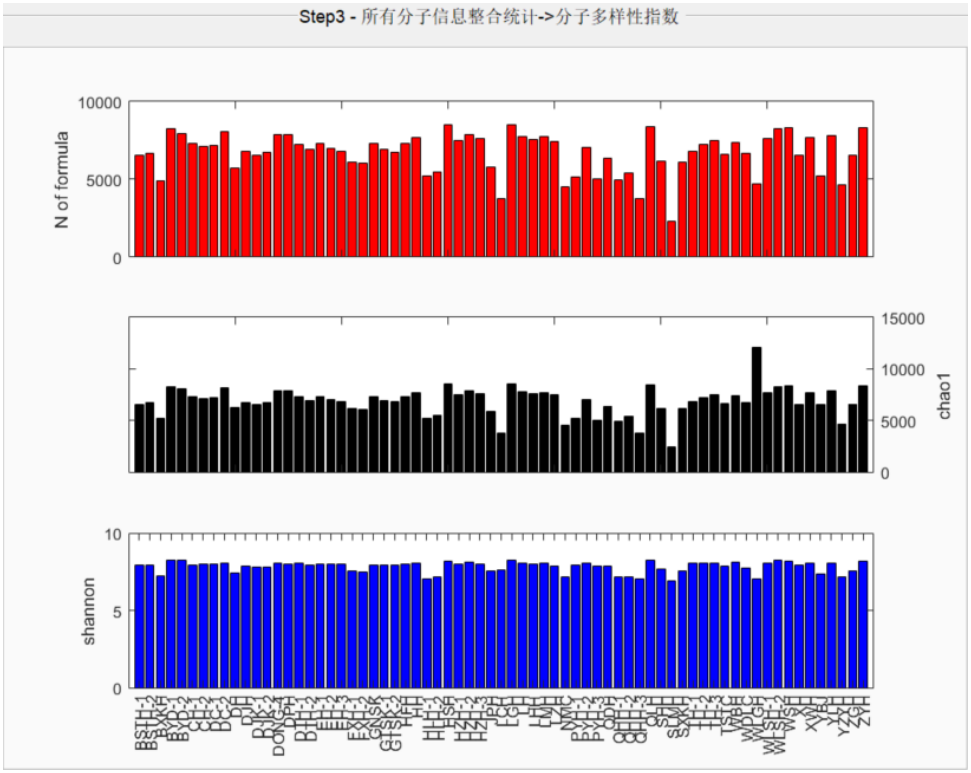


图 21 不同样品有机质分子多样性指数

3.3.11. 共同分子分析：第二级菜单选项包括共同分子数量与检出率关系、范氏图、分子特征、分子类型、聚类分析、样品间比较和特征关系等。

- 共同分子数量与检出率关系：根据 3.3.6 所述，共同分子与 DF 之间的关系将以点线图形式呈现（图 22）。

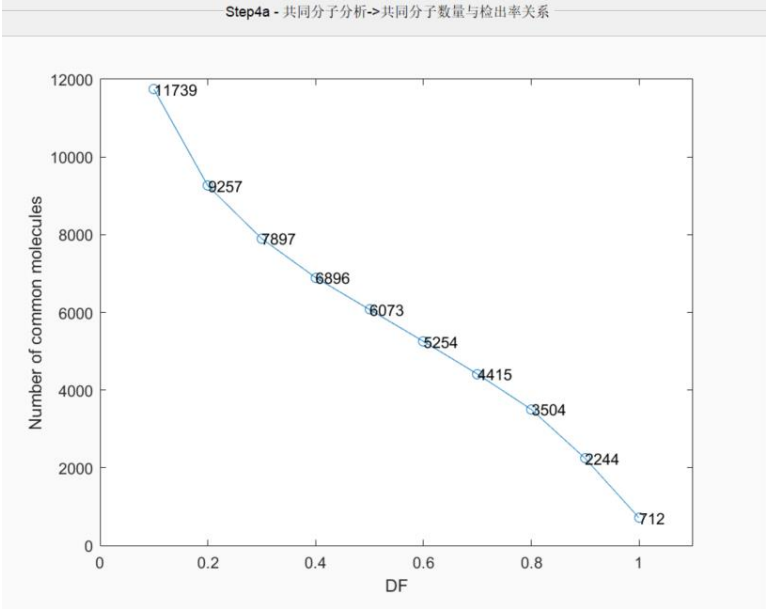


图 22 共同分子数量与检出率关系

- 范氏图：参考 3.3.9，样品菜单变为检出率菜单。
- 分子特征：参考 3.3.9，样品菜单变为检出率菜单。
- 分子类型：参考 3.3.9，样品菜单变为检出率菜单。
- 聚类分析：聚类分析主要采用的是层次聚类分析方法，聚类分析之前的样品间关系矩阵采用了不同的方法，本功能提供了四种聚类分析方法：共同分子强度欧式距离、共同分子强度相关系数、样品间共同分子出现率和共同分子强度热图+聚类（图 23）。其中，共同分子强度欧式距离是利用所有样品最大 DF 值情况下共同分子在各样品中的相对信号强度数据，采用欧式距离关系矩阵；共同分子强度相关系数是利用所有样品最大 DF 值情况下共同分子在各样品中的相对信号强度数据，计算样品间 Spearman 相关系数（r），利用 1-r 获得关系矩阵；样品间共同分子出现率，在 3.3.6 中计算得到了两个样品间共同分子在各自样品中的出现率矩阵（J），利用 1-J 获得关系矩阵；共同分子强度热图+聚类利用所有样品最大 DF 值情况下共同分子在各样品中的相对信号强度数据，制作不同各样品共同分子的热图，并在有机分子和样品两个维度上进行聚类分析，因有机分子数量庞大，计算较慢（注意：根据样品数量和计算机运算能力，可能需要等待较长时间），且该图不在本 GUI 界面显示。

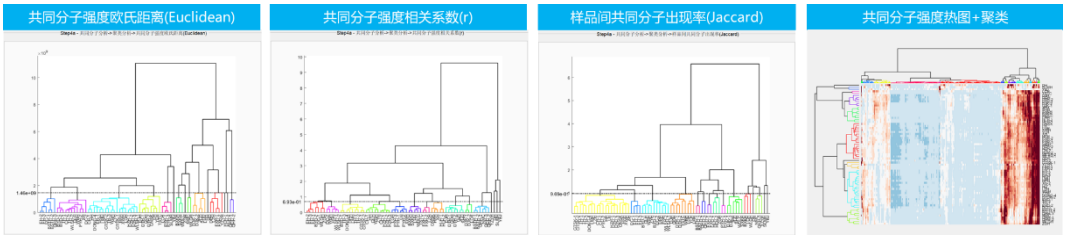


图 23 聚类分析图

- 样品间比较：该功能实现对两个样品间全部分子、共同分子、样品 1 独特分子、样品 2 独特分子以范氏图形式展示出现来（图 24）。

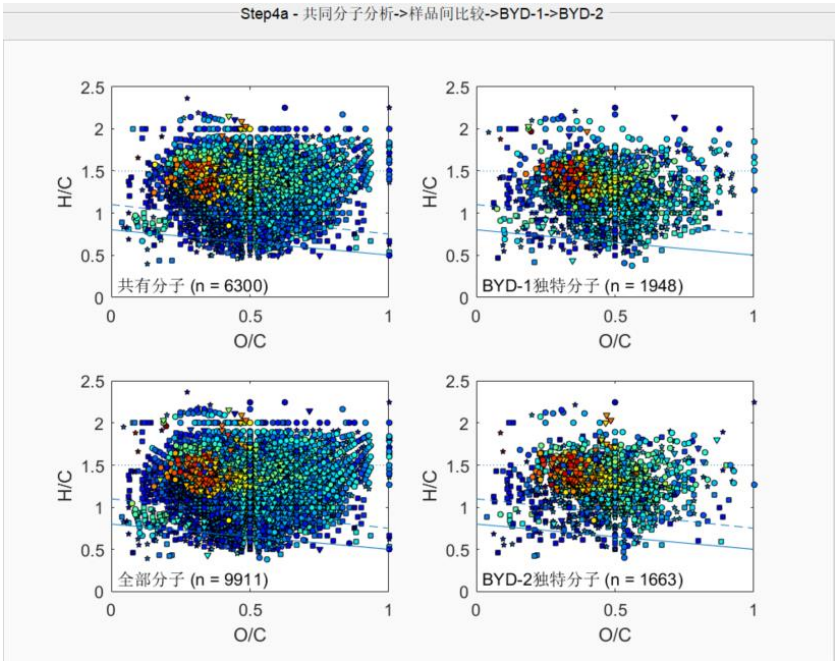


图 24 样品间共同和独特分子分析

- 特征关系：参考 3.3.9，样品菜单变为检出率菜单。

3.3.12. 独特分子分析：该功能是对仅出现在某个样品中的有机分子进行分析，包括范氏图、分子特征、分子类型和特征关系的显示。

- 范氏图：参考 3.3.9。
- 分子特征：参考 3.3.9。
- 分子类型：参考 3.3.9。
- 特征关系：参考 3.3.9。

3.3.13. 范氏图自定义绘制：点击菜单“作图”，选择“范氏图”，打开“绘

制范氏图”界面（图 25）。通过第一个“导入”按钮将拟作图的数据导入到软件中，数据整理要求见图 25，第一列、第二列和第三列分别为 O/C、H/C 和分子信息指标（见 3.3.4）。范氏图中每个点的颜色梯度依据信息指标数据。滚动条和后面的数字的调整可以修改范氏图散点的大小，实际操作中可以通过滚动条将散点调整到合适的大小，记下滚动条后的数字，以后的图可以通过输入对应的数字绘制统一点大小的图。点击“更新”可以更新图示，点击“保存”将保存当前图片到指定的文件夹中，保存图片的格式 3.3.8 所述，图片名称为 VK+时间数字。

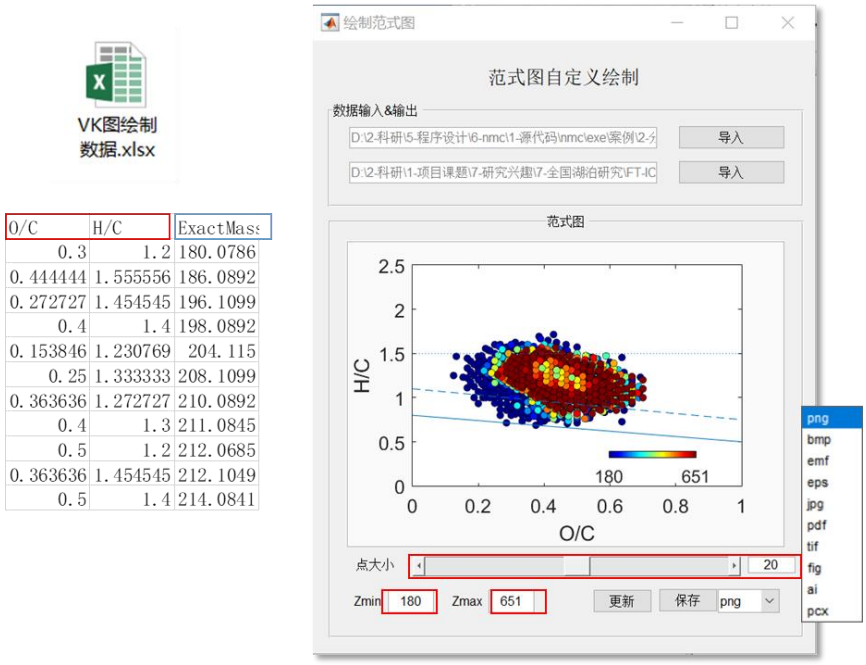


图 25 绘制范氏图界面

3.3.14. 分子网络界面：该界面包括“数据输入、输出、计算和保存设置”和“分子特征图示”两个部分（图 26），前者主要是实现数据输入与输出、结果数据的导入、样品内分子自关联网络分析、样品间共同分子自关联网络分析、以及样品间分子转化潜力网络分析；后者是前者 Step2 至 Step4 结果的展示，点击“更新”按钮可以获得相应的结果图。本功能对分子的网络分析基于两个分子之间存在基团（CH_2、CO_2、H_2、H_2O、O、SO_2、NH_2、C6H5）上的类似性，如，C2H6 和 C4H10 之间因相差 2 个 CH2 基团而存在关联。该界面保存的数据以 MAT 格式存在，可以通过 Step1.2 读取该数据；针对某个样品或者样品间分子网络可以通过 Step2 至 Step4 中的“保存当前”或“保存全部”来取得结果数据和图示。

- 样品内分子自关联网络分析：针对某个样品，计算分子之间是否存在上述 8 个基团的差异、以及所包含的 8 个基团的数量。基于分子之间的关联可构建分子网络，并得到不同基团的分子网络图。形成分子网络后，可以开展网络分析，主要得到分子网络的度（Degree，与某个分子有联系的分子数量）、度中心性（Degree centrality，体现分子核心性（重要性），该值越大说明该分子越重要，属于较活跃的分子，反之，该分子较不活跃）、亲密中心性（Closeness centrality，体现分子影响力，该值

越大说明该分子在整个样品中的辐射能力约强，即，除了与近端分子有直接联系，还可能与远端分子存在间接联系）、间接中心性（Betweenness centrality，体现分子作为中间过渡性分子的重要性，数值越大说明该分子在整个网络中具有较强的连接作用）、特征向量中心性（Eigenvector centrality，体现分子与邻近核心分子的关系，数值较大说明该分子与很多核心分子存在关系，该分子属于核心分子的联系或过渡分子）。分子网络构建完成后，可以点击“更新”得到分子网络图（图 27）。

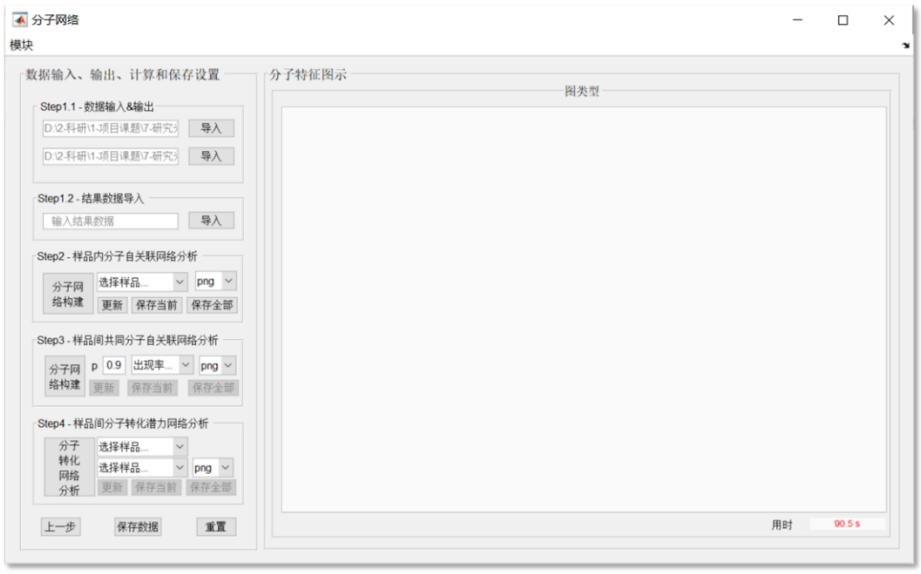


图 26 分子网络界面

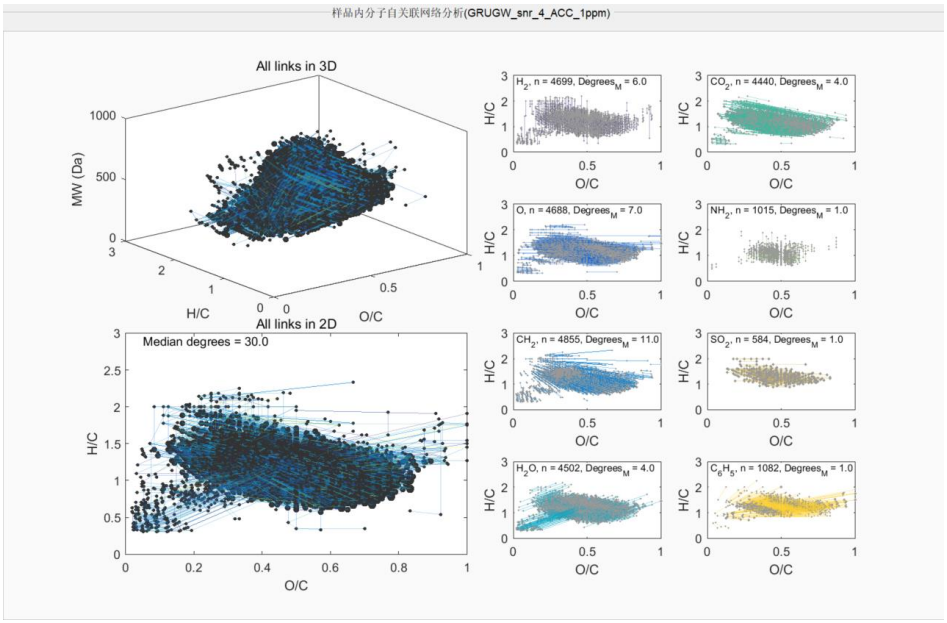


图 27 样品内分子自关联分子网络图

- 样品间共同分子自关联网络分析：采用 3.3.11 输出的不同检出（出现）率的共同分子数据集（这里排除出现率最低的一个共同分子数据集，以加快运行速率）；后续数据分析同“样品内分子自关联网络分析”。

- 样品间分子转化潜力网络分析：与“样品内分子自关联网络分析”中分子之间的关联无方向性不同，“样品间分子转化潜力网络分析”的分子关系是有方向性的，如，样品 1 分子 A 指向样品 2 分子 B，相差的基团是 CH2 的 n 倍，表明样品 1 分子 A 通过脱除 n 个 CH2 基团降解为样品 2 分子 B，换句话说，样品 2 分子 B 通过加成 n 个 CH2 基团形成了样品 1 分子 A，本网络图中的箭头正方向指向脱除过程，而反方向指向加成过程。此外，本功能只针对样品 1 中与样品 2 分子不重合的唯一分子转化与样品 2 中分子的相互转化潜力开展网络分析。针对具有方向性的分子网络分析，度和亲密中心性均包含入和出两个指标；此外还包括网页排名（Page rank，体现分子的重要性，数值越大说明该分子是被转化或者转化形成可能性最大的分子）、枢纽（Hubs，体现为样品中分子的被转化概率大小，数值越大说明分子被转化的可能性越高）、权威（Authorities，体现为样品中分子的形成概率大小，数值越大说明该分子越易由其他分子转化而成）。

3.3.15. 分子转化界面：该界面包括“数据输入、输出、计算和保存设置”、“样品间分子转化分析”和“样品清单”三个部分（图 28）。第一部分主要是分子信息模块结果数据输入、批处理表单提交和数据存储路径设置；导入数据后，会在第三个部分显示所有样品及其编号；在第二部分，可以选择被转化和转化后的样品，可以进行单独分子转化分析，也可以根据第三部分的样品清单在分子转化批运行表格中编辑被转化和转化后的序号，从而实现批处理分析。本软件转化算法依据的是 Danczak et al. (2020)[Danczak, R.E., Chu, R.K., Fansler, S.J. et al. Using metacommunity ecology to understand environmental metabolomes. Nat Commun 11, 6369 (2020). https://doi.org/10.1038/s41467-020-19989-y]提供的数据库，剔除涉 Cl 和 F 的转化过程，保留的转化过程共计 1148 种，可以归为 N 转化、S 转化、P 转化、NS 转化、NP 转化、SP 转化、NSP 转化和非 NSP 转化共 8 个大类。主要的输出结果包括每个分子涉及转化过程的个数、按照单个转化过程或转化大类的分子信息统计、用于绘制**桑基图**展示转化路径（按照元素、特征 I 和特征 II 分类）相关数据。

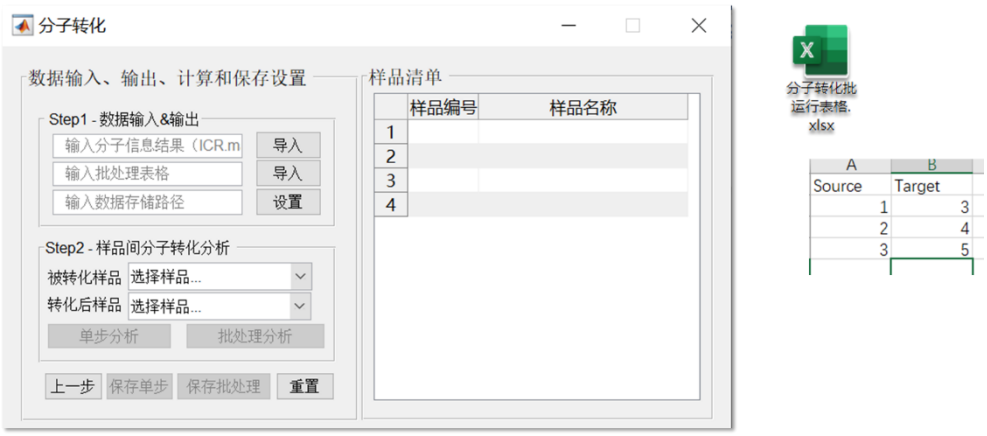


图 28 分子转化界面与批运行表格格式

3.4 分子结构界面

3.4.1. 分子结构界面分为“数据输入、输出、计算和保存设置”和“分子相关参数表格”两个部分（图 29），前者包括“数据输入&输出”设置和“选择查询分子”两个功能，后者以表格形式显示查询分子式潜在的结构参数。

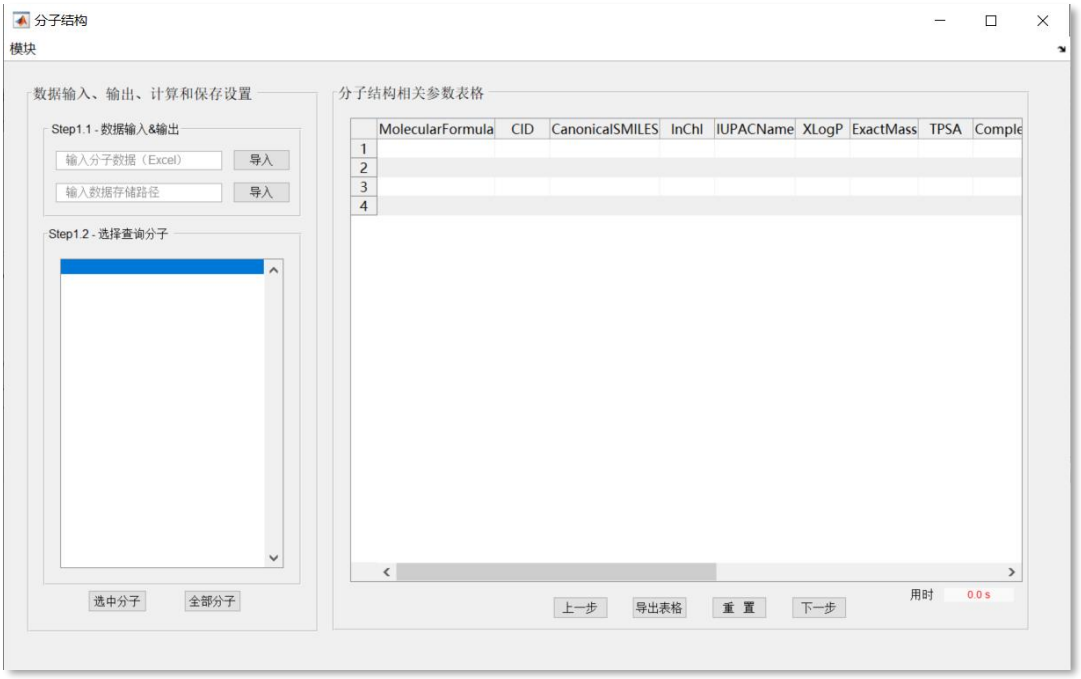


图 29 分子结构界面

当打开该界面时，软件会首先检查本电脑是否能访问 PubChem 数据库，是否能访问该数据库均会给出提示，如下图所示。

3.4.2. 数据输入&输出：导入数据需要按照图 13 生成的最大检出限共同分子数据集的形式进行整理，数据可存入 Excel 文档（xlsx 或 xls）和逗号分隔值文档（csv）中；数据输出的路径选择同 3.2.5。

3.4.3. 选择查询分子：导入数据后，所有分子式就会在清单表格中显示，如果只做某个分子的结构查询，在清单表中选中后点击“选中分子”（图 30），如果要做所有分子的结构查询，则点击“全部分子”，**由于每个分子的潜在结构数量较大，进行全部分子结构查询时需要等待较长的时间，期间可能因为数据库中断而发生中断；因此，运行全部分子时务必保证网络良好**。

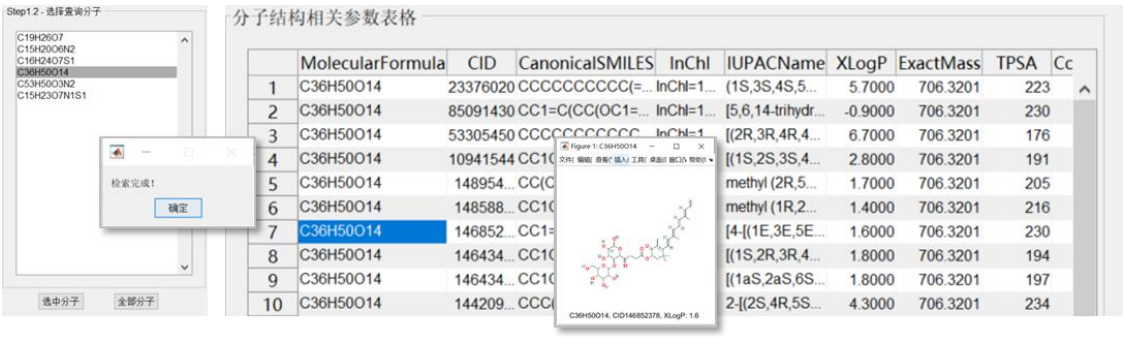


图 30 选中分子查询成功提示与结果

3.4.4. 分子结构相关参数：查询成功后，结构相关参数会以表格形式呈现，各列的英文代码及中文解释见表 1。点击“导出表格”后，会在结果文件夹中生成 PubChemResults+时间的 Excel 文档和 EPI_SMILES+时间的 TXT 文档，后者主要包含了 CID 和 SMILES 信息，用于后续模块。如图 30 所示，当选中某行时，软件会绘制该行分子结构图；通过“编辑”—“复制图形”可以拷贝该图片。

表 1 分子结构相关参数

英文代码	中文解释
MolecularFormula	分子式
CID	PubChem 化学品 ID
CanonicalSMILES	典型分子线性输入规范（SMILES）
InChI	国际化合物标识
IUPACName	国际纯粹与应用化学联合会命名
XLogP	辛醇水分布系数，疏水性测度
ExactMass	理论分子质量
TPSA	拓扑极性表面积
Complexity	分子复杂性评分
Charge	分子电荷
HBondDonorCount	氢键供体数目
HBondAcceptorCount	氢键受体数目
RotatableBondCount	可旋转键数目
Fingerprint2D	64 位编码的 PubChem 分子结构指纹

3.4.5. 导入结果数据：点击菜单“拓展”，选择“导入结果数据”，在弹出的对话框中选中 3.4.4 所生产的 Excel 文件，则可查看表中每个分子结构图。

3.4.6. 本地数据库：在查询分子时，软件均会首先从本地数据库中查询相关分子；若在本地数据库中存在该分子信息，则会直接读取数据库信息；若在本地库中不存在相关分子，则从 Pubchem 数据库中查询该分子，查询后会保存查询分子进入到本地库中。

3.4.7. 提示：**筛选查询分子时，建议一次性不要选中太多分子，由于 PubChem 数据库有海量的分子式，大量检索数据且长期以固定 IP 连接数据库，可能会让系统认定为机器爬取数据库，最终导致断开与用户 IP 的连接。**这里检索的相关分子一定是分子信息模块中发现的在信号强度和独特性方面具有重要特点。

3.5 活性信息界面

3.5.1. 活性信息界面分为“数据输入&输出”、“分子性质 VK 图”和“分子活性参数原始数据和统计分析”三个部分（图 31）。运行本界面前需要先安装 EPI suite，并记住所安装的路径，一般默认安装路径为 C:\EPISUITE41\。本软件内置了一个分子活性信息数据库，且通过“入库”按钮可以将每次非库中数据导入到数据库，从而丰富数据库。

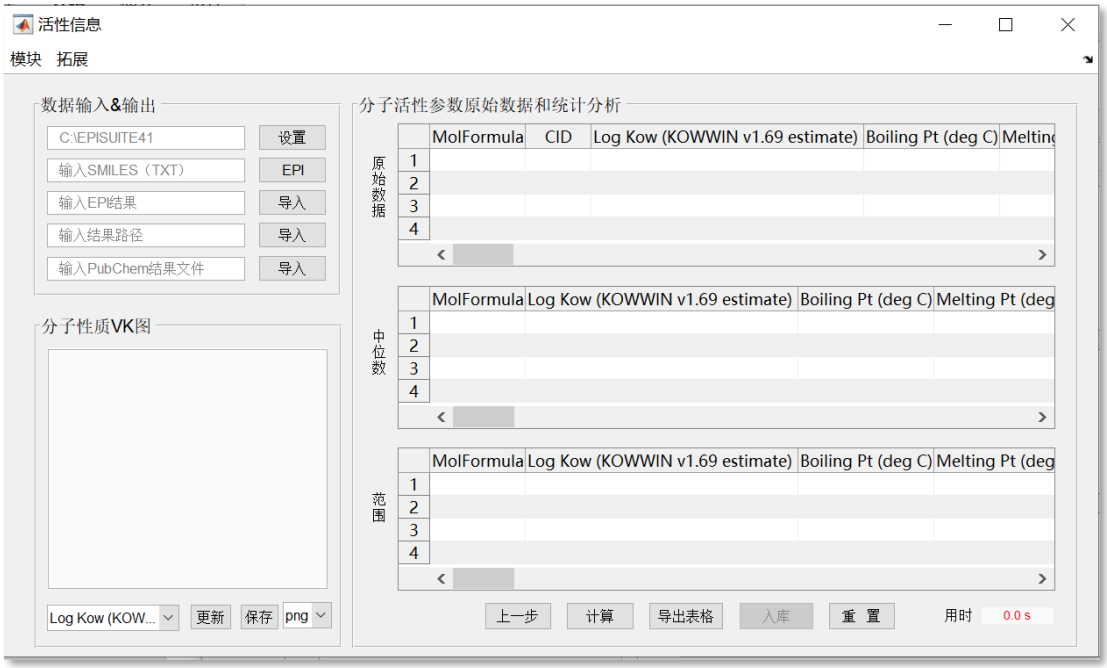


图 31 活性信息界面

3.5.2. 数据输入&输出：当 EPIsuite 未安装在默认路径时，先进行“设置”；随后，点击“EPI”选择 3.4.4 所保存的 SMILES 数据集，如果所有 SMILES 在本地数据库中可查询，会出现提示，告知直接从数据库读取，否则，会打开 EPI suite 的批运行功能（注意：为了激活该功能，需要先正常打开 EPIweb 界面，在 Batch Mode 中打开 Batch File Input Using Strings F5，如果依然无法调出批运行界面，可多打开几次），点击“StringFormat”按键，在对话框中选择刚才导入的 SMILES 数据集（该数据集中如果有部分 SMILES 存在于数据库中，则会被自动剔除，以减少 EPIsuite 计算工作量），而后，弹出名称为“Mark Batch Options, Then Start”的对话框，选择“Summary Output”，点击“Start Batch”；运行结束后，在对话框右上部分会显示结果名称（如图 32 中“EPIBAT01.OUT”）。随后，“导

入”EPI suite 输出文件（EPIBAT03.OUT），按照 3.2.5 介绍选择数据输出路径。最后，“导入”连接文档，即上一模块输出的 Excel 结果文件。

3.5.3. 分子性质范氏图：这里能将不同有机分子活性参数以范氏图的形式展示出来，横纵坐标为有机分子的 O/C 和 H/C 值，散点颜色梯度指示活性参数的大小变化，通过菜单可以选择需要显示的参数，相关参数英文名称及和中文解释见表 2。在菜单序列中选中某个参数指标时，点击“更新”可以显示该图示，点击“保存”可以保存该图，保存的图示类型包括矢量图（pdf、fig、eps 和 svg）和位图（png、jpeg、tiff 和 bmp）。

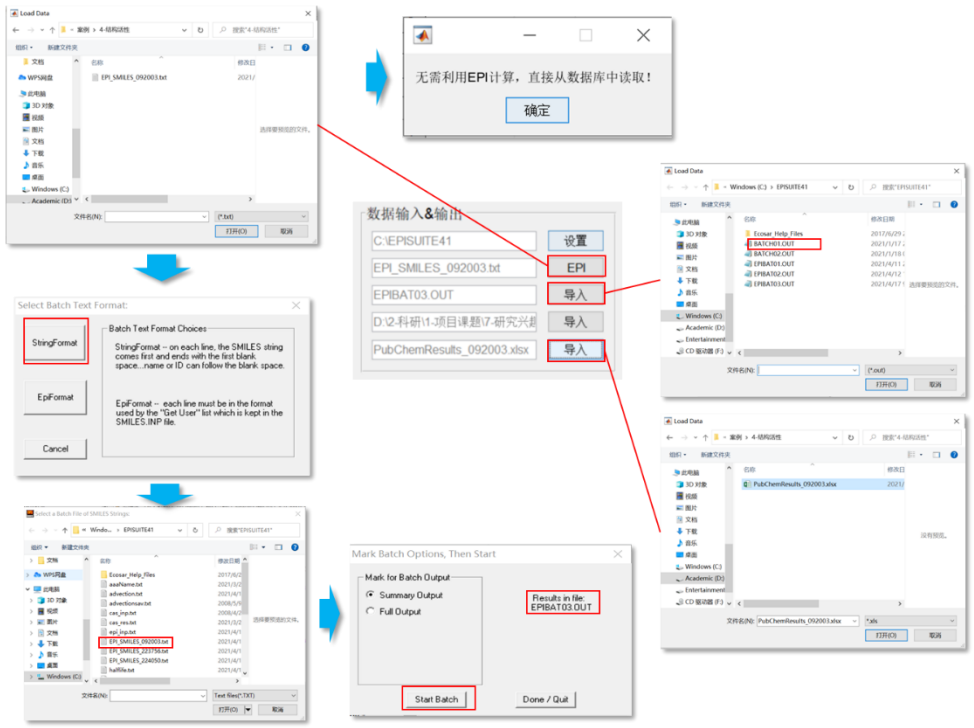


图 32 数据输入&输出的功能介绍

表 2 分子结构活性参数

英文名称	中文解释
Mol Formula	分子式
Log Kow (KOWWIN v1.68 estimate)	辛醇水分布系数
Boiling Pt (deg C)	沸点
Melting Pt (deg C)	熔点
VP (mm Hg,25 deg C)	蒸气压
VP (Pa, 25 deg C)	蒸气压
Subcooled liquid VP	过冷液体蒸气压
Water Solubility at 25 deg C (mg/L)	水溶解性
Wat Sol (v1.01 est)	水溶解性
Log Koa (KOAWIN v1.10 estimate)	正辛醇空气分配系数

Biowin1 (Linear Model)	线性生物降解概率模型
Biowin2 (Non-Linear Model)	非线性生物降解概率模型
Biowin3 (Ultimate Survey Model)	专家调查最终生物降解模型
Biowin4 (Primary Survey Model)	专家调查初级生物降解模型
Biowin5 (MITI Linear Model)	日本国际贸易产业省线性生物降解模型
Biowin6 (MITI Non-Linear Model)	日本国际贸易产业省非线性生物降解模型
Biowin7 (Anaerobic Linear Model)	厌氧生物降解模型
Kb Half-Life at pH 7	碱性水解半衰期
Log BCF from regression-based method	生物富集因子
Log Biotransformation Half-life (HL)	生物传递半衰期
Half-Life from Model River	模式河流半衰期
Half-Life from Model Lake	模式湖泊半衰期
Total removal	污水厂处理比率
Persistence Time	三级逸度模型评估的持久性
OVERALL OH Rate Constant (SingleOr[Cis])	大气羟基自由基反应速率（单分子或顺式结构）
OVERALL OH Rate Constant (SingleOr[[Trans])	大气羟基自由基反应速率（单分子或反式结构）
OVERALL Ozone Rate Constant (SingleOr[Cis])	大气臭氧反应速率（单分子或顺式结构）
OVERALL Ozone Rate Constant (SingleOr[[Trans])	大气臭氧反应速率（单分子或反式结构）
LogKoc (MCI method)	土壤吸附系数（MCI 法）
LogKoc (Kow method)	土壤吸附系数（Kow 法）

3.5.4. 分子活性参数原始数据和统计分析：点击“计算”，就能在这个功能界面得到 EPIsuite 计算的活性参数的结果(原始数据、中位数、最大值和最小值)。点击“导出表格”后，会在结果文件夹中生成 PropertyResults+时间的 Excel 文档，该文档中包含“原始数据”、“中位数”、“最小值”和“最大值”等 4 个 Sheet。

3.5.5. 导入结果数据：点击菜单“拓展”，选择“导入结果数据”，在弹出的对话框中选中 3.5.4 所生产的 Excel 文件，则可查看表中每个分子结构图以及进行范氏图的绘制和输出。

3.6 工作路径和案例数据

在 nmc 软件可执行程序的主路径中会有“案例”和“结果”等文件夹（图 33），其中“案例”数据包含了“操作方法”部分所需的案例数据；“结果”为数据结果所在的文件夹，实际操作时，可在界面中自定义存储用户数据结果的文件夹，此外，还包括一些数据库文件。

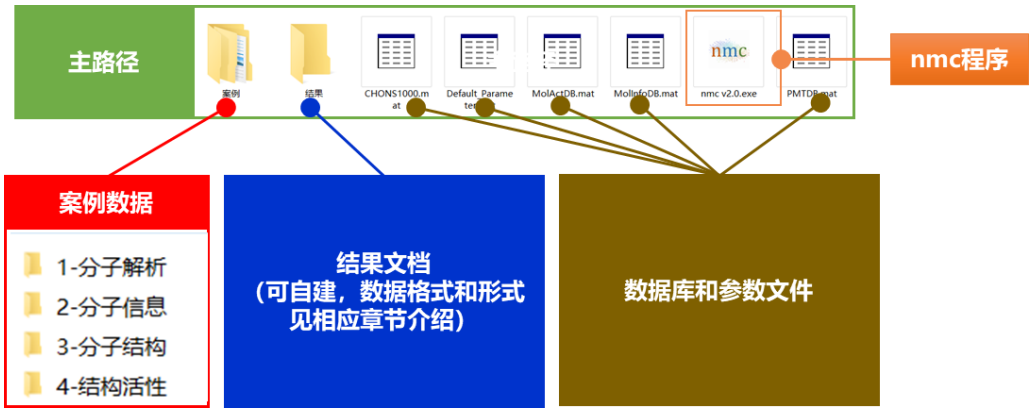


图 33 工作路径与案例数据

4. 操作方法

4.1 软件安装

本软件包含有可执行程序 exe 版本和 pcode 版本。exe 版本无需计算机有 Matlab 软件，但需要安装 Matlab Compiler Runtime，由于本软件采用 Matlab 2018b 开发完成，需要安装相应的编译器，即 Matlab Compiler Runtime 9.5。pcode 版本则需要完整的 Matlab 软件。

4.1.1 exe 版本

压缩包解压后，则会出现图 33 所示的工作路径；双击 nmc v2.0.exe，会出现图 2 的主功能列表界面。如果长期（5 min 以上）未出现该界面，检查是否安装了正确的编译器，也可与开发者联系，告知运行环境。

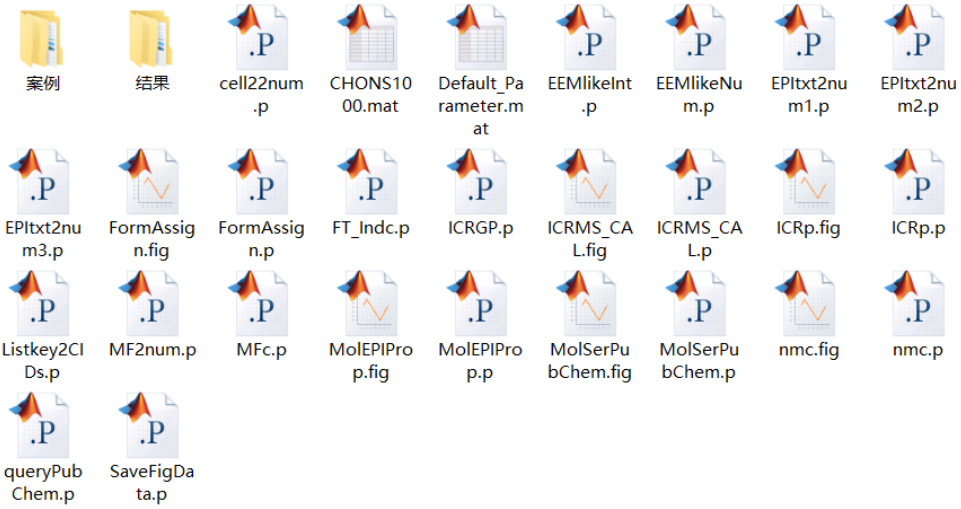


图 34 pcode 文件解压及目录下文件

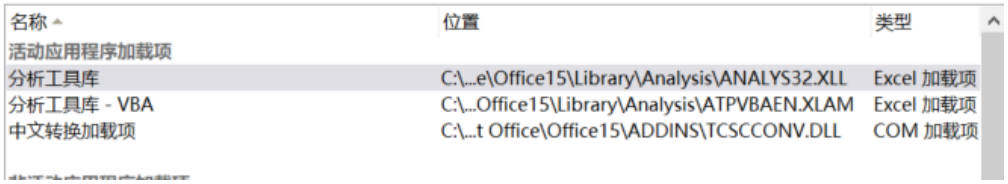
4.1.2 pcode 版本

对于熟悉 Matlab 编程的用户来，本软件也提供 Matlab 的 p 文件代码，解压

压缩包后如图 34 所示。首先，将 pcode 文件夹拷贝到 Matlab 的 toolbox 目录下；然后，利用 Matlab 的“File”里的“set path”将 pcode 文件夹及其子文件夹设置为工作路径即可；最后，在 Matlab 的命令窗口（Command Window）输入 nmc 即可打开主功能列表界面。

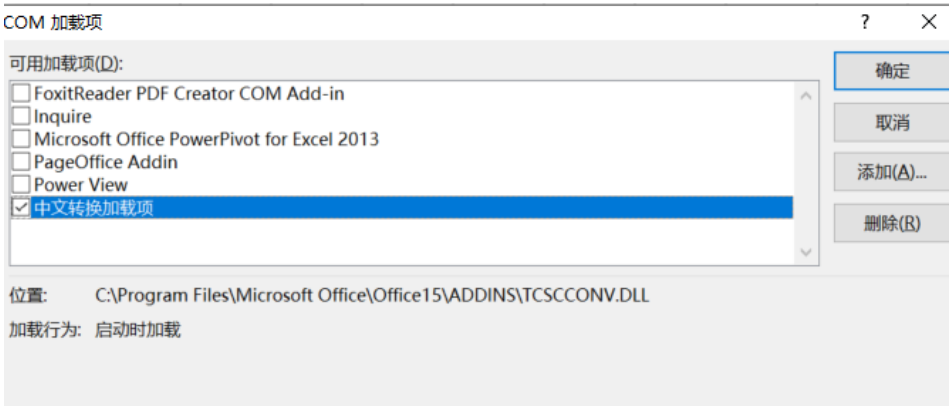
4.1.3 其他注意事项

由于软件需要调用微软的 Microsoft Excel COM 端口用于生成 Excel 等文档，因此，务必保证电脑中安装有 Microsoft 公司的 Office 64 位软件，建议版本为 Office 2010 以上。另外，打开 Excel，在“文件”—>“选项”—> “加载项”中保证如下图所示的“活动应用程序加载项”中类型为“COM 加载项”的应用不多于 2 个。



如果超过 2 个，请在界面左下角管理中选择“COM 加载项”。

点击“转到(G)…”，进入 COM 加载项对话框，通过勾选，取消掉多于的 COM 加载项，点击确定即可。



4.2 有机质的分子解析运行

针对分子解析界面的使用，本软件提供了案例数据，该数据在“案例”-“1-分子解析”文件夹中。如图 35 所示，一般操作流程是：

（1）设置分子数据库的元素组成及相应的最大数量，本案例设置采用软件默认值；

（2）设置同位素，本案例采用默认值；

（3）设置元素组成，本案例选择了“CHO”、“CHN”、“CHON”、“CHOS”和“CHONS”五类基于元素组成的化合物类别；

（4）设置质量范围，本案例采用默认值；

（5）点击第一个“导入”；

（6）打开“案例”文件夹，进入“1-分子解析”文件夹，按住 Ctrl，多选三个 Excel 文件，单击“打开”；

（7）点击第二个“导入”；

（8）选择结果保存文件夹，不选择任何文件夹，直接点击“选择文件夹”按钮；

（9）设置精确度（ppm）为 1、信噪比（SNR）为 4 和 DBE 最大值为 20；

（6）点击“计算”，进行分子解析；

数据文件夹在输出路径里可以找到，一般命名为：分子解析结果+时间。

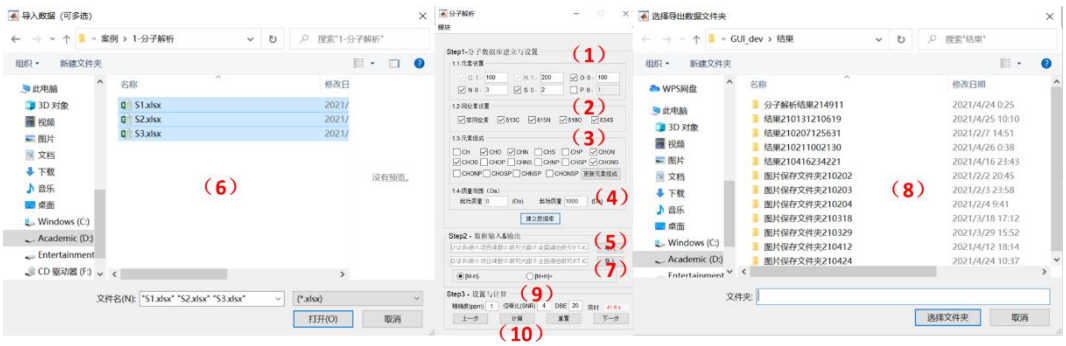


图 35 分子解析的一般操作流程

4.3 有机质的分子信息计算

针对分子信息提取界面的使用，本软件提供了案例数据，该数据在“案例”-“2-分子信息”文件夹中。如图 36 所示，一般操作流程是：

（1）点击第一个“导入”；

（2）打开“案例”文件夹，进入“2-分子信息”文件夹，选择 Excel 文件，单击“打开”；

（3）在弹出的参数设置对话框中，点击“关闭”，使用默认值；

（4）点击第二个“导入”，直接点击“选择文件夹”按钮；

（5）设置精确度（ppm）数值为 1，信噪比数值为 10，不勾选“+H”复选

框，单击“分子特征汇总和分类”按钮；

（6）选择 O/C 和 H/C 间距倍数均为 5，点击“运行”；

（7）选择共同分子出现率间隔为 0.2，点击“运行”；

（8）选择相关性分析为 Pearson，相关系数绝对值为 0.9999999，点击“运行”；

（9）点击“保存数据”；

（10）第一级菜单选择“Step4a – 共同分子分析”，第二级菜单选择“范氏图”，第三级菜单选择“NOSC”，第四级菜单选择“1”，点击“更新”；

（11）图片保存格式选择“png”，点击“保存”，弹出“保存图示文件”对话框，单击“当前图示”，等待图片保存完毕。

数据文件夹在输出路径里可以找到，一般命名为：结果+长时间数字。保存图片也可在输出路径相关文件夹中找到，一般命名为：图片保存文件夹+日期数字。一般性流程中保存的图片在图片保存文件夹+时间数字下的“Step4a - 共同分子分析”- 范氏图文件夹中，共有 5 个图片和一个 Excel 文件。

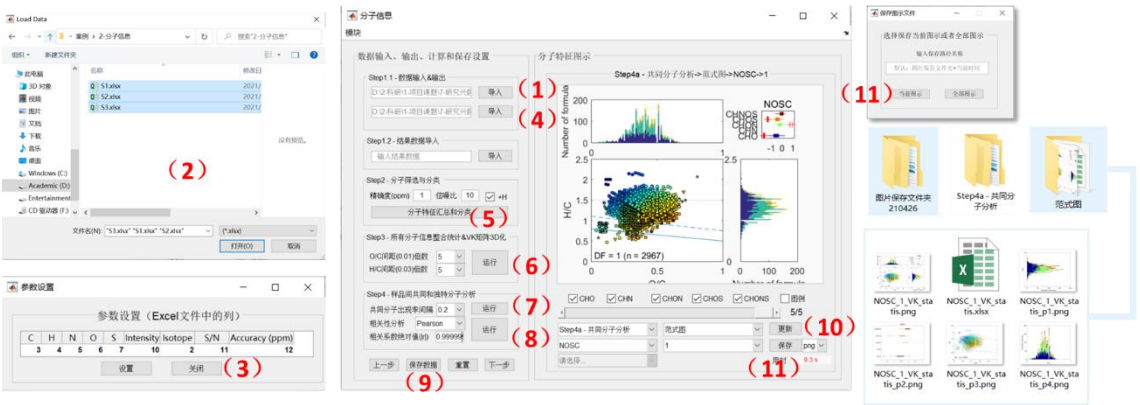


图 36 分子信息提取的一般性流程

4.3.1 案例 1：导入未处理样本并显示范氏图（3D）

按照如图 36 所示的一般性流程多选择案例文件中的三个 Excel 文档。其他步骤一样，在步骤（10）中第一级菜单选择“Step3 – 所有分子信息整合统计”，第二级菜单选择“范氏图（3D）”，第三级菜单选择“分子数量”，第四级菜单选择“S2”，点击“更新”，则会得到图 37 所示的等值线图。

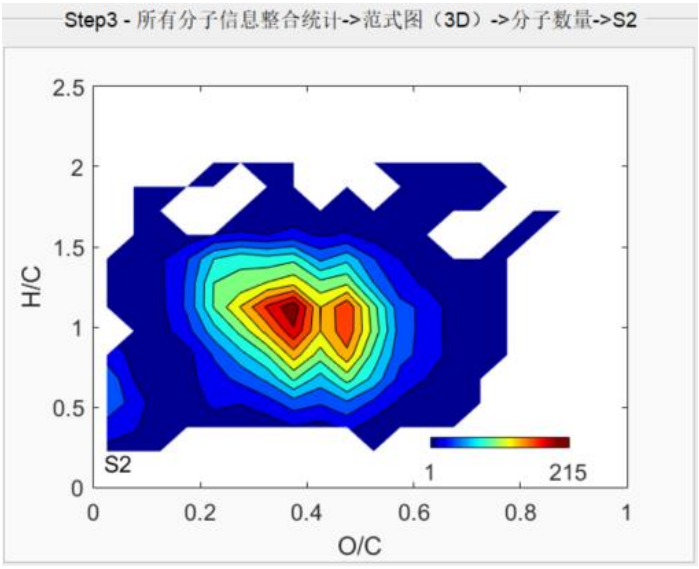


图 37 样品 S2 的 3D 范氏图

4.3.2 案例 2：导入结果数据并显示样本 S2 独特分子 KMD（CH_2）图

一般性流程中的步骤（1）和（2）替换为点击第三个“导入”，弹出文件选择对话框，在“案例”-“2-分子信息”文件夹中选择名称为“ICR(acc1sn10)130256”的 mat 数据文件，点击“打开”，等导入成功提示框出现后表明，数据导入成功；一般性流程步骤（10）第一级菜单选择“Step4b – 独特分子分析”，第二级菜单选择“特征关系”，第三级菜单选择“ExactMass”，第四级菜单选择“S2”，第五级菜单选择“CH2”，点击“更新”，则会得到图 38 所示的 KMD 图。

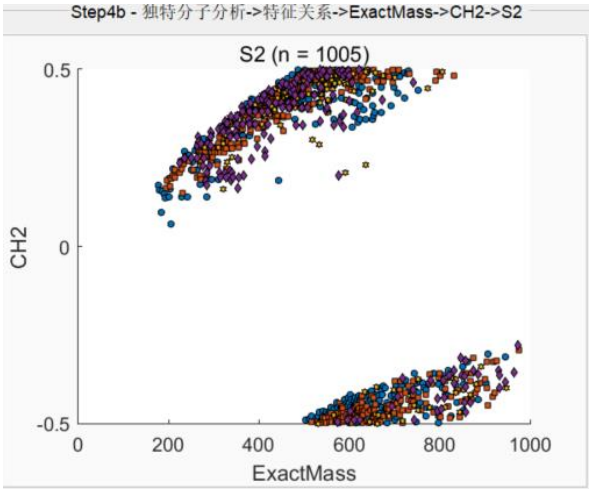


图 38 样品 S2 的 KMD（CH_2）图

4.4 有机质的分子网络分析

针对分子网络分析界面的使用，本软件提供了案例数据，该数据在“案例”-“3-分子信息”文件夹中。如图 39 所示，一般操作流程是：

（1）点击第一个“导入”；

（2）打开“案例”文件夹，进入“2-分子信息”文件夹，选择分子信息模块输出的 Mat 文件，名称为 ICR(acc1sn10)130256，单击“打开”；

（3）点击第二个“导入”，直接点击“选择文件夹”按钮；

（4）点击 Step2-样品内分子自关联网络分析中的“分子网络构建”按钮，等进度条结束关闭后，灰色的“更新”、“保存当前”和“保存全部”等按钮将可使用；选择样品 S1，点击“更新”按钮可在右侧分子特征图示中显示该样品内各分子之间的关联；

（5）点击 Step3-样品间共同分子自关联网络分析中的“分子网络构建”按钮，等进度条结束关闭后，灰色的“更新”、“保存当前”和“保存全部”等按钮将可使用；p 值设置为 0.9999，检出率选择 1，点击“更新”按钮可在右侧分子特征图示中显示该样品内各分子之间的关联；

（6）点击 Step4-样品间分子转化潜力网络分析中的“分子转化网络构建”按钮，等进度条结束关闭后，灰色的“更新”、“保存当前”和“保存全部”等按钮将可使用；源样品选择 S1，标的样品选择 S2，点击“更新”按钮可在右侧分子特征图示中显示该样品间各分子之间的转化关系；

（7）在 Step4-样品间分子转化潜力网络分析中选择“png”图片保存格式，保存当前图示，则在“结果”文件下的“网络分析图片保存（Step4）+时间”中可以找到相关图片以及网络分析数据；

（8）点击“保存数据”，在“结果”文件夹下的“分子连接结果+时间”文件夹中可以找到 Mat 格式的结果数据；

（9）点击第三个“导入”，选择（8）步骤输出的结果数据，则可以重复（4）至（6）中的“更新”、“保存当前”和“保存全部”等步骤。

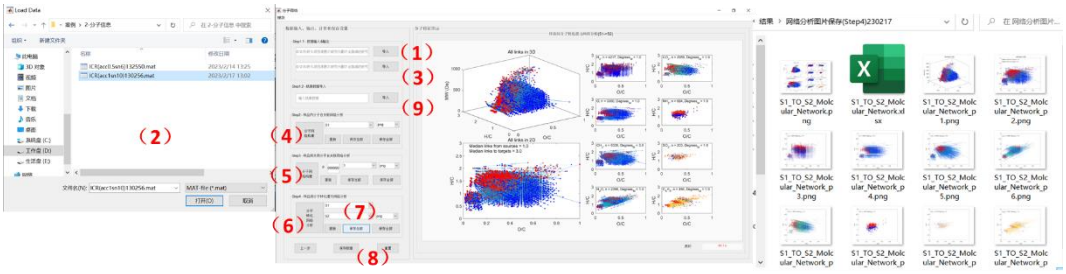


图 39 分子网络分析的一般性流程

4.5 有机质的分子转化分析

针对分子转化分析界面的使用，本软件提供了案例数据，该数据在“案例”-“3-分子信息”文件夹中。如图 40 所示，一般操作流程是：

（1）点击第一个“导入”；

（2）打开“案例”文件夹，进入“2-分子信息”文件夹，选择分子信息模

块输出的 Mat 文件，名称为 ICR(acc0.5sn6)1325506，单击“打开”；

（3）点击“设置”，而后点击“选择文件夹”按钮；

（4）选择被转化样品（来源）和转化后样品（靶标）后点击“单步分析”，由于转化机制较多，此处分析至少需要 10 min，待精度条结束后，分析完成；

（5）点击“保存单步”可将单步分析数据保存在 Excel（命名为样品 1_TO_样品 2_PMT）中；

（6）点击第二个“导入”；

（7）打开“案例”文件夹，进入“2-分子信息”文件夹，选择分子转化批运行表格，单击“打开”；

（8）点击“批处理分析”，软件按照批运行表格中序列进行分子间转化分析；

（9）点击“保存批处理”，可将批处理分析数据保存在多个 Excel（命名为样品 1_TO_样品 2_PMT）中。

图 40 分子转化分析的一般性流程

4.6 有机质的分子结构检索

针对分子结构检索界面的使用，本软件提供了案例数据，该数据在“案例”-“3-分子结构”文件夹中。如图 41 所示，一般操作流程是：

（1）点击第一个“导入”，弹出文件选择对话框，在上述“3-分子结构”文件中找到名称为“Common_molecule_select.xlsx”的 Excel 文档，选择后点击“打开”；

（2）点击第二个“导入”，直接点击“选择文件夹”按钮；

（3）点击“选中分子”或“全部分子”；

（4）点击右侧表格中的“132606…”；

（5）在弹出的分子式图片中菜单栏选择“编辑”-“保存图形”；

（6）点击“导出表格”。

可以在输出路径里找到 2 个数据文件，一般命名为：PubChemResults_时间数字.xlsx 和 EPI_SMILES_时间数字.TXT。

4.6.1 案例 1：选择单个分子

步骤（3）选择“选中分子”。

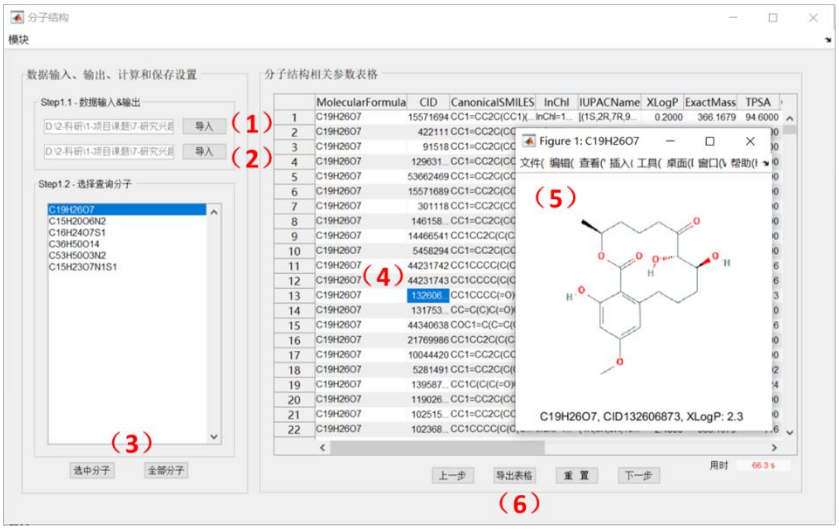


图 41 分子结构检索的一般性流程

4.6.2 案例 2：选择全部分子

步骤（3）选择“全部分子”。

4.7 有机质的分子活性计算

针对结构活性计算界面的使用，本软件提供了案例数据，该数据在“案例”-“4-结构活性”文件夹中。如图 42 所示，一般操作流程是：

（1）点击“设置”，在弹出的文件夹选择对话框中点击“选择文件夹”；

（2）点击“EPI”，在弹出的文件选择对话框中，选择案例数据中名称为 EPI_SMILES_092003.txt 的数据文件，点击“打开”，同时打开 EPI suite 的“Select Batch Text Format”界面；

（3）在“Select Batch Text Format”选择“StringFormat”，弹出“Mark Batch Options, Then Start”对话框；

（4）在“Mark Batch Options, Then Start”对话框中选择“Summary Output”，点击“Start Batch”，计算介绍后点击“Done/Quit”退出；

（5）点击第一个“导入”，在弹出的文件选择对话框中选择（4）步的计算结果文件 EPIBAT+数字.OUT ；

（6）点击第二个“导入”，直接点击“选择文件夹”按钮；

（7）点击第三个“导入”，选择上一模块产生的结果数据；

（8）单击“计算”；

（9）下拉菜单中选择“Kb Half-Life at pH 7”，点击“更新”；

（10）选择图片格式为 png，点击“保存”，在弹出的对话框中选择“当前图示”；

（11）点击“导出表格”。

数据文件在输出路径里可以找到，一般命名为：PropertyResults_时间数字.xlsx。保存图片也可在输出路径相关文件夹中找到，一般命名为：图片保存文件夹+日期数字，图片名称为 PubChemResults_时间数字_Kb Half-Life at pH 7.png。

图 42 分子结构活性计算的一般性流程