²H-SIAM(1.0) Algorithm and ²H-SIAM Pipeline

Introduction

²H-SIAM(V1.0) is a feature list-based algorithm developed for mapping possible transformation products (TPs) of stable isotope-labeled organic matters. With the help of this algorithm, we developed a ²H-labeled Stable Isotope Assisted Metabolomics (²H-SIAM) pipeline which facilitate the mapping of TPs with high confidence. Additionally, our proposed ²H-SIAM pipeline is equipped with full GUI interface which significantly lower the skill barriers of its application in research field.

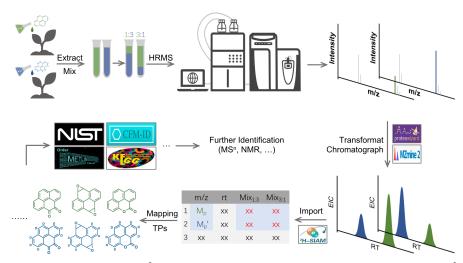


Figure 1. Proposed pipeline for ²H-labeled Stable Isotope Assisted Metabolomics (²H-SIAM). Firstly, both natural and ²H-labeled parent contaminants are individually incubated in environment matrix for indicated duration; they are subsequentially extracted and mixed with the ratio of 1:3 and 3:1 (Mix_{1:3} and Mix_{3:1} samples); the mixtures are determined by LC-ESI-HRMS or GC-EI-HRMS and thereafter obtained raw data are analyzed by MZmine2 to obtain feature list; then it is imported into our developed algorithm ²H-SIAM(V1.0) to obtain possible isotopologue pairs as transformation products of contaminants for further identification.

Important tips: Although the proposed ²H-SIAM pipeline is designed for the mapping of ²H-labled TPs by UPLC-ESI-HRMS, data from any other types of isotope-labeled atoms, e.g., ¹³C, ¹⁵N and ¹⁸O, or any other types of LC-MS or GC-MS platform, or any other mass spectrum data processing algorithm, e.g., XCMS, MZmine2, MetaboAnalyst and MS-DIAL, are compatible for our proposed ²H-SIAM pipeline.

²H-SIAM pipeline

As shown in Fig. 1, a fully GUI-based pipeline is proposed, and it facilitate to lower the skill barriers of the application of ²H-SIAM in environment study. Parent isotopologue pairs of contaminants are added into environmental matrix

respectively and incubated for indicated duration. Then, they are extracted and mixed with indicated ratio (e.g., 1:1 and 1:2, 1:3 and 3:1, 1:9 and 9:1). Here we propose a 1:3 and 3:1 ratio (Mix_{1:3} and Mix_{3:1} samples) which keep an appropriate equilibrium between accuracy and sensitivity of the mapping of isotopologue pairs.

Mixed samples are then determined by MS platform (e.g., UPLC-ESI-HRMS, GC-MS) for indicated replicates. MS raw data of Mix_{1:3} and Mix_{3:1} are transformed into .xzXML format (by ProteoWizard 3.0.20353) and imported into MZmine2 to obtain feature list information. Then, Mix_{1:3} and Mix_{3:1} data in feature list was imported into ²H-SIAM(V1.0) for TPs mapping.

Any other mass spectrum data processing algorithm, e.g., XCMS, MetaboAnalyst and MS-DIAL, are compatible if they could output feature list information as .csv format with m/z, rt and intensity (height/area). Here we propose peak height, rather than peak area, as quantitative filters, as chromatographic trailing in HPLC may significantly affect reproductivity of peak area. The key for efficient mapping TPs labeled by isotopes is to acquire high quality feature list, and then optimized parameters for the selected algorithm are necessary.

How to use ²H-SIAM(V1.0)

²H-SIAM(V1.0) accept data from any mass spectrum data processing algorithm, e.g., XCMS, MZmine2, MetaboAnalyst and MS-DIAL, in a .csv format. A feature list which is ready for use in ²H-SIAM(V1.0) is shown in Fig. 2.

2 60 3 60 4 6 5 63	0.0534014 0.2603882 61.029299 3.7893214 4.0166295	9.58938719 15.4015725 6.3897052 12.7523426	4.74179362 3.32718716 25.5668419	6.45391255 3.77633901	9.49996235	8.15031894			Mix3v1−1 P∈	Mix3v1-2 P€	Mix3v1-3 Pe	Mix3v1-4 Pe	Miv3v1_5 Pe	Mix3v1_6 Pe	ak height
3 60 4 6 5 63	0.2603882 61.029299 3.7893214 4.0166295	15.4015725 6.3897052 12.7523426	3.32718716 25.5668419	3.77633901			0	0					INIIVOAT OLG	INIIVOAT OLC	an incignit
4 6 5 63	61.029299 3.7893214 4.0166295	6.3897052 12.7523426	25.5668419		4.86463585			U	8.23417798	0	11.8223004	7.96992713	0	5.32542415	
5 63	3.7893214 4.0166295	12.7523426		10 2200765		2.3667658	0	7.10431636	5.11390236	2.62822296	2.58237039	5.95388082	0	4.39094929	
	4.0166295		2.40104261	10.2303103	32.7835548	24.7423277	35.1519265	18.3642296	20.210568	18.3737346	14.9066701	10.9908504	6.33057595	3.19988118	
6 64		3 9286561	3.49104201	3.25257416	4.08821879	3.79601607	0	3.45806987	3.59996878	3.26550682	5.15946469	4.40171513	4.23488707	2.37686791	
	5.0268543	0.0200001	3.24509871	4.05959006	5.32629956	3.59582403	5.21662408	4.92668156	4.44052733	4.69736776	3.91524268	0	3.4622314	2.92382202	
7 65		26.0290098	9.10849282	8.76738026	0	9.81963026	16.481173	0	10.8663454	11.6007272	0	6.39194742	0	0	
8 65	5.0390549	14.653747	16.3708069	17.4432704	31.9776666	19.7125718	15.9246425	16.1614012	12.2313499	13.0067821	8.46969196	8.71387247	5.90039477	8.75594725	
9 68	8.0332565	19.8194984	6.1079396	9.7014614	11.6081521	9.91854263	0	10.6025977	0	6.65669658	6.68535866	0	3.84885641	0	
.0 68	8.0641015	21.0101205	3.07815414	4.39381677	6.61407531	4.55310059	4.86219397	4.93400814	6.3109003	7.02843879	5.9900682	6.13532276	5.33166292	2.89130879	
.1 68	8.3367406	6.95026493	184.919848	246.527241	389.999782	261.096265	61.5431583	75.1056291	18.1710971	0	6.73683125	17.372745	19.3350137	7.05936103	
		6.26617811	0			-				33.9946219	5.9994591	22.5034447	13.8021604	16.0088448	
3 68	8.9979421	1.13131778	51.3359072	43.8468367	56.5358873	52.9944694	62.1072811	55.4435314	0	33.4513921	48.91871	37.3718904	22.7702258	26.4016273	
		6.25608828								574.266183	512.088207	315.671314	164.669901	163.474783	
.5 68	8.9979425	7.06572336	902.876888	299.144464	376.899736	719.246023	525.515257	351.866852	390.888373	0	314.892122	281.001662	164.208013	0	
6 69	9.0340218	21.7224951	35.902876	40.7026724	41.8424024	0	70.6125889	50.1389654	0	11.7382363	11.136997	12.2113835	8.66440001	7.97376675	
7 69	9.0343103	13.0533758	17.9398018	18.5126207	6.22039344	18.0291045	5.66915458	13.5340592		18.1524166	3.40474652	13.3360139	0	10.7124771	
.8 69	9.0343179	11.9688248	95.3658494	133.600998	178.797392	155.506709	151.212655	149.946671	123.5861	126.08219	122.603182	113.054558	93.2380044	70.7422858	
	9.0706916				105.321461					56.8792154					
		8.61008399			14.7193416							10.7625409		7.97902683	
21 69	9.0706937	13.8483623	0	19.5139671	23.8329695	16.0110454	17.5748395	0	14.9891484	20.9648189	19.8120742	13.7650327	12.1187309	16.9954338	
										17.9897642					
-					5.44274968						7.51806046		-	3.66991543	
24 70	0.0132496	1.12579333		0	-		42.6554493						119.746004	0	
	0.0488815		10.2123799		19.0695526				0				8.22693152		
			8.49738144			9.26208015		16.1927019	0				6.22004574	5.49947389	
-		25.0407292			19.1828534						0		5.38844555	0	
28 70	0.0659313	27.9491221			437.464952		9.10960211		0	0	0		·	Ü	
		4.17494903			12.6312442					18.3142992		7.79820754		0	
30 70	0.0669285	26.4826223	8.49846829			14.8869939	15.0059572			11.1598864	7.50022466	9.92222666			
			3.76556464			5.79128876		0	0	0		5.23329649		3.17446603	

Figure 2. Typical feature list in a .csv format used for ²H-SIAM(V1.0).

The first row is titles, including m/z, RT and intensity of Mixture1 and Mixture2 with indicated replicates (defined by customs). The first and second columns define m/z and RT (minutes) of features obtained from your data processing algorithm. Important Notes: the sequence of the feature list should be sort by

the increasement of m/z as shown in Fig. 2!!! The following columns are intensities (height/area) of mixture samples (e.g. Mix_{1:3} and Mix_{3:1} samples) measured by LC-MS or GC-MS. Replicates could be defined by customs.

After preparing your feature list, running ²H-SIAM(V1.0) and the following GUI interface will appear (Fig. 3). A typical setting for mapping potential TPs of pyrene by Mix_{1:3} and Mix_{3:1} samples by UPLC-ESI-HRMS (Q Exactive, Thermo) is as following: Mass tolerance 7 ppm, RT tolerance 1 min, Number of labeled atoms 3-10, Mass difference of atoms between labeled or not (Da) 1.006174, Ratio for F1, F2 and F3, 0.3333, 3 and 3, Tolerance for F1, F2 and F3, 0.3, 0.3 and 0.5.

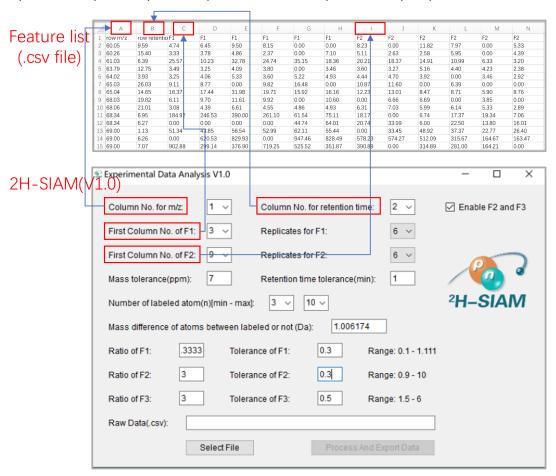


Figure 3. GUI interface of ²H-SIAM(V1.0) and its correlation with feature list (.csv).

Then select your feature list file and process, which will take several minutes depending on your PC and number of features. Our example feature list contain 4407 features, and it takes 1 minutes for TPs mapping in a PC equipped with 3.7GHz Intel Core i9. Additionally, if you have more than 10000 features, time may be significant prolonged, e.g., ~20 min. After that, a table will appear, and you could double click the table and save it as a .csv file.

When feature list (.csv format) is imported into 2 H-SIAM(V1.0), it is start with qualitatively searching of isotopologue pairs with indicated isotope label, m/z and RT tolerance (Fig. 4, Step 1), and then 3 quantitative filters F_1 , F_2 and F_3 are used to select potential natural TPs with a isotopologue (Step 2-4). Finally, as shown in

Fig. 5, $^2\text{H-SIAM}(V1.0)$ will output a feature list .csv file (Fig. 4) and those features as potential TPs will be marked with numbers of potential $^2\text{H-labels}$. In some experiment regime, you may wish use just filter F1, and then you could remove tick before "Enable F1 and F2".

Ste	p 1: Pa	aris Sea	ırch		Step 2: Filter 1 (0.333, 0.11-1.1)									
			Mix _{1:3}	Mix _{3:1}				Mix _{1:3}	Mix _{3:1}					
1	M_p	XX	XX	XX	1	M_p	XX	XX	XX					
2	M _p '	xx	XX	XX	2	_M _p '	XX	XX	XX					
3	XX	XX	XX	XX	3	xx	xx	XX	XX					
Ste	p 4: Fil	ter <mark>3</mark> (3	, 1.5-6.0		Ste	p 3: Fil	ter <mark>2</mark> (3	, 0.9-10)						
	m/z	rt	Mix _{1:3}	Mix _{3:1}		m/z	rt	Mix _{1:3}	Mix _{3:1}					
1	M_p	XX	XX	xx	1	M_p	XX	XX	XX					
2	M _p '	XX	XX	XX	2	M _p '	xx	XX	XX					
3	XX	XX	XX	XX	3	XX	XX	XX	XX					

Figure 4 Algorithm for ^2H-SIAM. When feature list (.csv format) is imported into 2 H-SIAM(V1.0), it is start with qualitatively searching of isotopologue pairs with indicated isotope label, m/z and RT tolerance, and then 3 quantitative filters F_1 , F_2 and F_3 are used to select potential natural TPs with a isotopologue.

	A	В	C	D	E	F	G	H	- 1	J	K	L	M	N	0	P	Q	R		S	T U	V	W
m/z		Retention Ti			A3	A4									Mean of F1								
																					.: 1.60266069597564;		
		12.1680182																			.: 1.82376905965496;		
202	077737																				.: 2.63855290407218;		
	0.05575		5.07562871		8.63566126			9.05645921													.: 3.2072244925924;		
																					.: 3.77328013188073;		
		13.0080351																			: 1.53662347245175;		
																					: 1.55003346622176;		
																					: 1.68109105876558;		
																					: 1.73506823860279;		
																					: 2.23462176436988; 7	.: 1.7469452	2571083
																					: 1.52014282266633;		
																					: 1.59241010525153; 6	.: 1.8700280	3869971
																					: 1.86469090506496;		
																					2.05661274022646;		
		1.21430385																			: 2.65995415030585;		
																					: 2.66522559988572;		
			28.0327611																		: 3.08100079159331;		
		13.8444041	0																		: 1.67563764892399;		
		1.14585897		14.1353679																	2.47934883441864;		
		28.5205467		18.5541265				33.7998284													2.67637505251117;		
																					: 4.0915115347696; 7L	: 3.47927796	5167529
		14.882402					23.9104444				19.6616888		15.8645271								: 1.52190624417104;		
		13.7338965				3.6106073															: 1.70896556223601;		
		12.7586157																			: 1.73410875321473;		
		28.541974				12.6808203		27.5507214													: 1.77627167518016;		
	075396 133968																				1.86307111500999;		
																					2.05236143735226;		
		27.9926361	0 0545750		7.46114731																: 2.1183182784549; : 2.30171896458755;		
196	077541	28.545217	28.9015/59	37.9731861	00.8522077	20.1422432	35.9621956	56.0020524	Z8:303Z367	ST:2880812	37.4123742	14.4762153	23.4438261	17.1419259	41.5239102	23.8404449	/L@/193.1	TT VF : 0.93	roam (F :)	1.2103B /L	: Z.3U1/1896458755;		

Figure 4 Typical output .csv document list isotopologue pairs as TPs. Only data that pass indicated filters will be marked with the format "nL: XXXX", where n refers number of labeled atoms and XXX refers to the calculated ratio.

Data availability

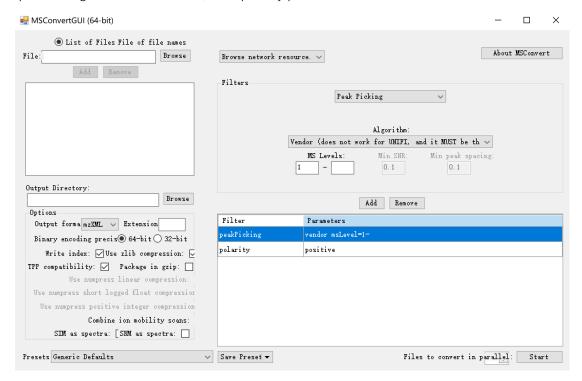
We offer our example data for practice of ²H-SIAM pipeline. Following data could be found from our website.

18 .raw file of the experiment
FeatureListOutputFromMZmine2.csv
PyreneSoilFeatureListImportFor2HSIAM.csv
OutputFrom2H-SIAM.csv
BatchModelForC18PyrSoil1T3_3T1.xml

ProteoWizard (3.0.20353) for windows MZmine2 (2.53) for windows ²H-SIAM(V1.0) for windows

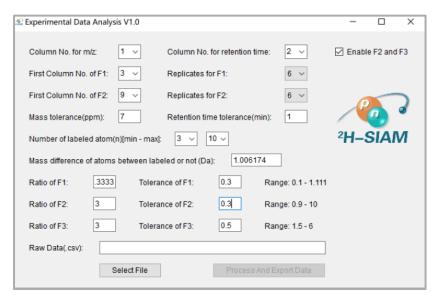
Description of the data:

18 .raw file of the experiment, contain raw data from UPLC-ESI-HRMS (Q Exactive, Thermo) including blank, Mix1:3 and Mix3:1 with 6 replicates. It should be format to xzXML file by ProteoWizard (3.0.20353, or other version) with the following setting: Output format mzXML, Binary encoding precise 64-bit, Write index yes, Use zilib compression yes, TPP compatibility yes, Filter peakPicking vendor msLevel=1-, Filter polarity positive.



FeatureListOutputFromMZmine2.csv, feature list obtained by MZmine2 by batch model using the script offered here. Standard compound normalizer and export to .csv is complished manually. Before use, blank data should be deleteded and data should be sort by the increasement of m/z. An example is uploaded with the file name PyreneSoilFeatureListImportFor2HSIAM.csv

PyreneSoilFeatureListImportFor2HSIAM.csv, typical ready for use .csv file for 2H-SIAM. **OutputFrom2H-SIAM.csv**, 2H-SIAM(V1.0) analysis output file of "PyreneSoilFeatureListImportFor2HSIAM.csv" with the following setting:



BatchModelForC18PyrSoil1T3_3T1.xml, MZmine2 analysis script for example data. Additionally, Standard compound normalizer was accomplished manually as following setting: Nearest standard, Peak measurement type peak height, m/z vs RT balance 1, Standard compounds m/z 188.1405 at 16.14 (anthracene-d10). Then, .csv could be exported with m/z, rt and peak height.

ProteoWizard (3.0.20353), windows software for MS raw data format.

MZmine2 (2.53), windows software for MS data analysis.

2H-SIAM(V1.0), windows software for isotope-label metabolomics.

Description of the example experiment:

pyrene and pyrene-d10 were dissolved in acetone (ACE) and mixed respectively with one quarter of blank soil. After solvent evaporation, they were mixed with the rest unspiked soils to obtain 100 ppm pyrene and pyrene-d10 contaminated soil respectively. Then, blank and contaminated soils were placed in room temperature for 60 days and watered weekly to keep wet. After 60 days incubation, soil samples from blank, pyrene and pyrene-d10 treatment were harvested and air dried at 60 °C until to constant weight. 1 gram of contaminated soils were extracted by 10 mL of 1:3 acetone and Hexane by microwave extraction, and anhydrous NaSO₄ were used to remove residual water. After that solvent was replaced by acetonitrile by solvent exchange, and extracts were concentrated to 1 mL under nitrogen flow.

Extracts from pyrene and pyrene-d10 were mixed with the ratio of 1:3 and 3:1 to obtain Mix_{1:3} and Mix_{3:1} samples. Anthracene-d10 were added into blank and mixed samples as internal standard. Then, blank and mixed samples were subjected to UPLC-ESI-HRMS analysis (6 replicates).

 $5~\mu L$ of samples was injected into UPLC-ESI-HRMS system. UPLC solvents were A, water in 0.1% formic acid, and B, acetonitrile with 0.1% formic acid. UPLC were performed at 1 mL/min at 25 °C with the following linear gradient (minutes, %B): 0, 5%; 4, 5%; 8, 95%; 26, 95%; 28, 5%; 30, 5%.

Mass spectrometry was detected via profile mode (scan range, m/z 60-900) with a resolving power of 140 000 fwhm (at m/z 200) and automatic gain control setting of 3×10^6 with a maximum injection time of 200 ms. The heated electrospray ionization (ESI) source was operated using the following settings: sheath gas flow rate, 40 au; auxiliary gas flow rate, 20 au; spray voltage, 3.8 kV; capillary temperature, 325 °C.