

pirika.comの研究記録

Pirika News



科学は嘘をつかない。統計は嘘をつく。

YMB24Pro4MI用のプログラム作成

大学の授業が終わり、さて何をしようか考えた。とりあえず、久しぶりに本格的なプログラミングかな。

というわけで、長年の課題、HSP用の距離の式に取り組んだ。HSPiPには搭載されていないので一般のユーザー向けの話ではない。YMB24Pro4MIを利用可能なユーザー以外は今月号はスキップして欲しい。

HSPiPの δH の分割

HSPiPの水素結合項 δH は、2010年に販売開始のver 3.1から δH_{acid} , δH_{base} に分割された。HSPiPでは δHD , δHA , ドナー/アクセプターと表記されている。しかし、あくまでもプロトンのドナーとプロトンのアクセプターなので、酸/塩基の表記の方が正しい。

この分割を行うために、UCL(University College London)のAbraham教授を訪ね、さまざまな化合物のAcid/Baseの値を頂いた。そしてSmilesの構造式からAbraham Acid/Baseを予測するQSAR式を構築した。つまり、SMILESの構造式があればAbraham Acid/Baseは予測できる。

そして、


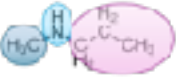


$$dH_{acid}:dH_{base} = \text{Abraham-acid}:\text{Abraham-base}$$

$$dH^2 = dH_{acid}^2 + dH_{base}^2$$

の関係になるように dH_{acid} と dH_{base} を決定した。活性水素を持たないものは、基本的には dH_{acid} はゼロになる。そこで、ほとんどの化合物の dH は dH_{base} に割り振られる。

基本HSPが同じであれば同じ溶解性？

これは、HSP 50周年記念講演会(2017年)の時に私が発した疑問。

		δH
中性		[15.9, 4.5, 5.9] MVol=131.9, BP=399.26K 5.9
塩基性		[15.6, 3.9, 5.8] MVol=100.9, BP=336.15K 5.8
酸性		[16.2, 3.5, 6.9] MVol=241.4, BP=585.25K 6.9
中性		[16.1, 3.0, 6.1] MVol=323.3, BP=624K 5.1

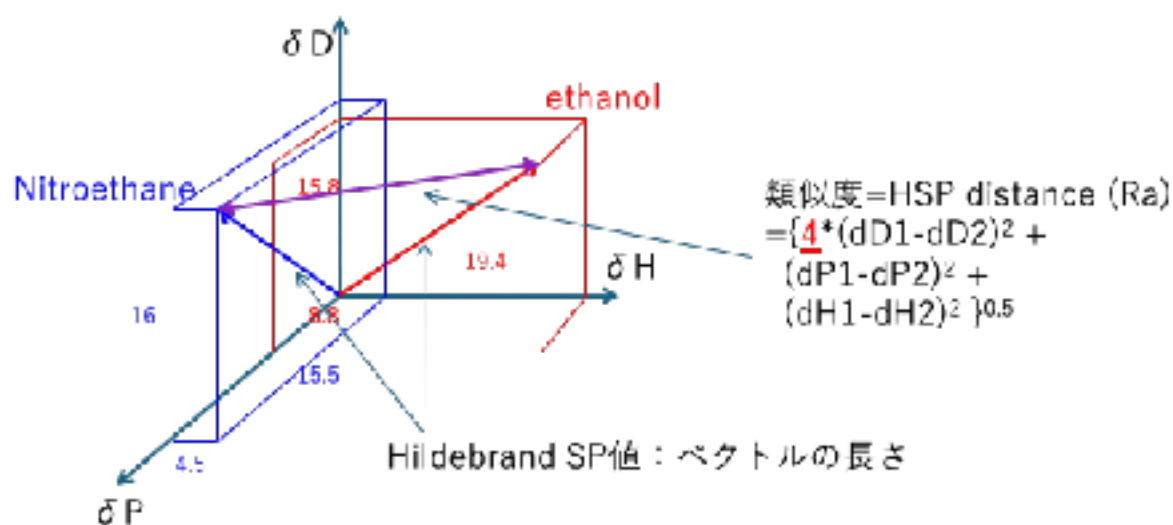
$$\sigma = \{(\Delta H - RT)/V\}^{0.5}$$

HSPのトータル値は蒸発潜熱(ΔH)と分子体積(V)から求まる。すると、ほぼ同じHSP値を持つ化合物は分子体積次第で先の図にあるように多数存在する。

HSPの基本概念は、HSPが似ていれば溶解性が似ているとなる。

しかし、中性、酸性、塩基性の化合物の溶解性が同じになるとは化学者なら誰も信じないだろう。

HSP値の類似度



HSPの類似性を計算する時には、HSP距離(Ra)を使う。dDの前につく4.0というファクターに関しては常に論争的になった。その解釈については後に述べる。

しかし、問題点は、一番よく溶解するのはHSP距離がゼロとなる溶質と溶媒のHSPが完全に一致した時であるという事。つまり塩基性顔料とか酸性吸着剤だとかの分散に溶媒の酸性、塩基性、中性は関係なくなってしまう。

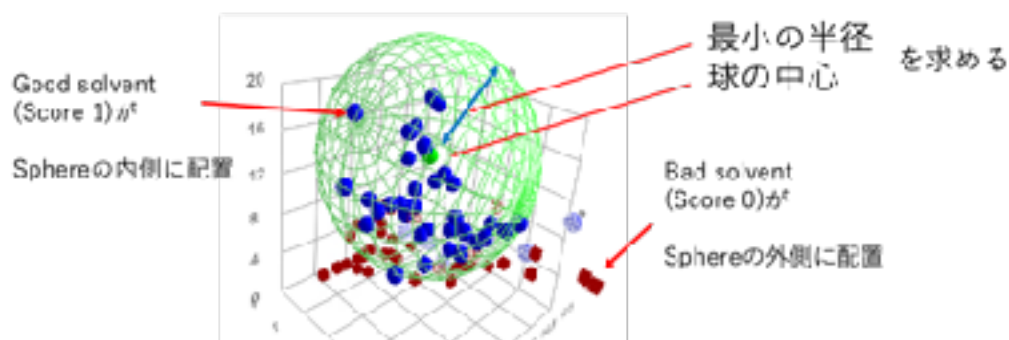
新たに、dHacidとdHbaseを定義したらHSP距離(Ra)をどう取るかが2010年に問題になった。最初にAbbott教授がver. 3.1に搭載したのは、dHacidとdHbaseによって計算値を変えるとんでもない式だった。HSP距離が短い方がより良い溶解性とは言えない代物だった。

ver.いつから切り替わったのか知らないが、最終的にはEuclid Typeの次式が搭載され現在に至っている。

$$\sqrt{4.0*(dD1-dD2)^2+(dP1-dP2)^2+(dHacid1-dHacid2)^2+(dHbase1-dHbase2)^2}$$

この式の問題点は、HSPのオリジナルと同じで、酸塩基の相互作用は加味されないことだ。

ハンセンの溶解球とは？



Good solventがハンセンの溶解球の内側に、Bad Solventが溶解球の外側に配置されるような緑色のメッシュの球を求めるのがハンセン先生のSphere法だ。

その時の球の中心が溶質のHSP値となる。

ある未測定の前媒のHSPの3次元上の位置がハンセンの溶解球の内側にくる場合、研究者が定義した、Goodに相当するくらいその溶質を溶かすことが期待できる。半径を長くすればWrong Inという間違って入ってくるものが増える。半径を短くとるとWrong Outという間違って出してしまうものが増える。そのバランスをハンセン先生はFit値として評価する。

バランスを取っても Wrong In/Outはどうしても現れる。

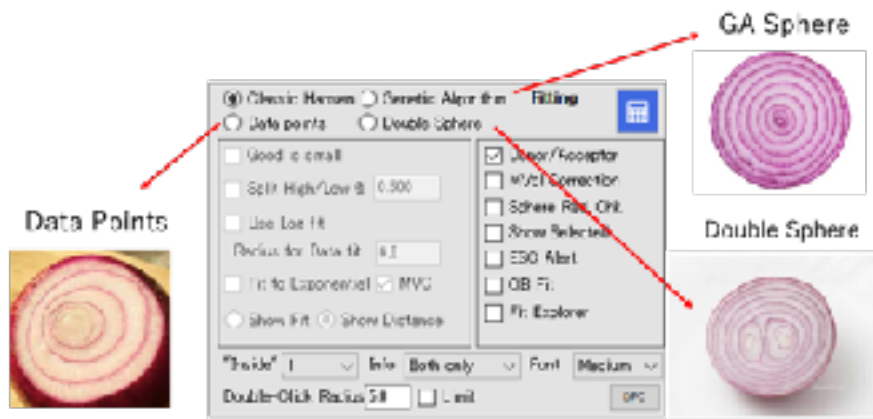
Wrong Inの解釈は、例えばポリマーの可塑剤のように、HSP的にはポリマーと非常に近いのだが、分子量が大きいのので浸透できず溶解できないことがある。

Wrong Outの解釈は、小さな分子がエントロピー的に溶解するとかも言われている。

正確にいうと、境界領域で、そんなに精度の出る方法ではないという事だ。期待しすぎないでほしい。

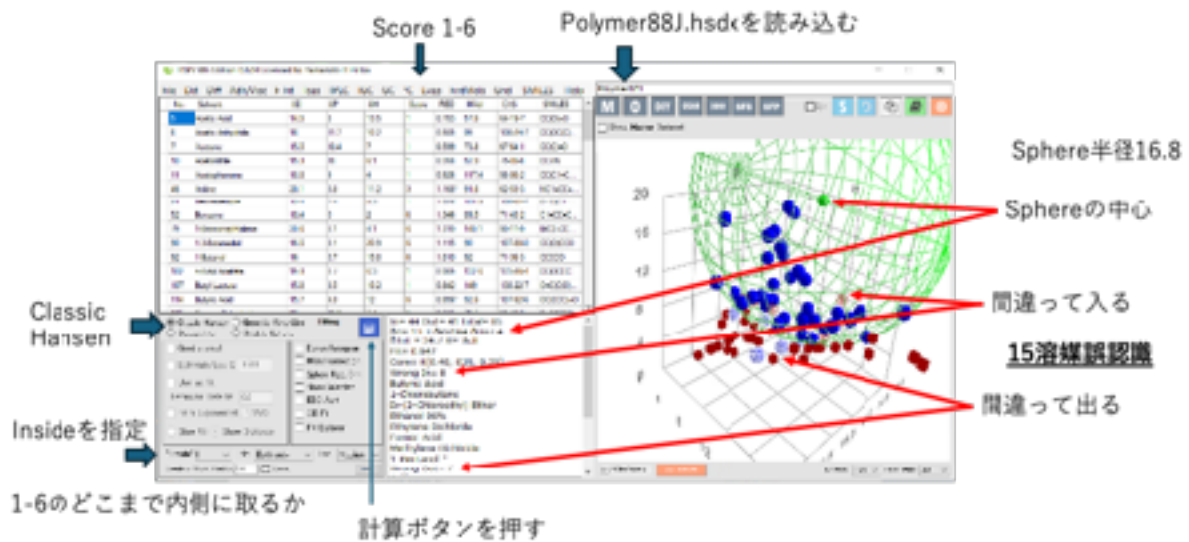
ハンセンの溶解球を求める4つの方法

HSPiPには溶解球を求める4つの方法が搭載されている。元のClassic Hansen以外は私の作った方法だ。



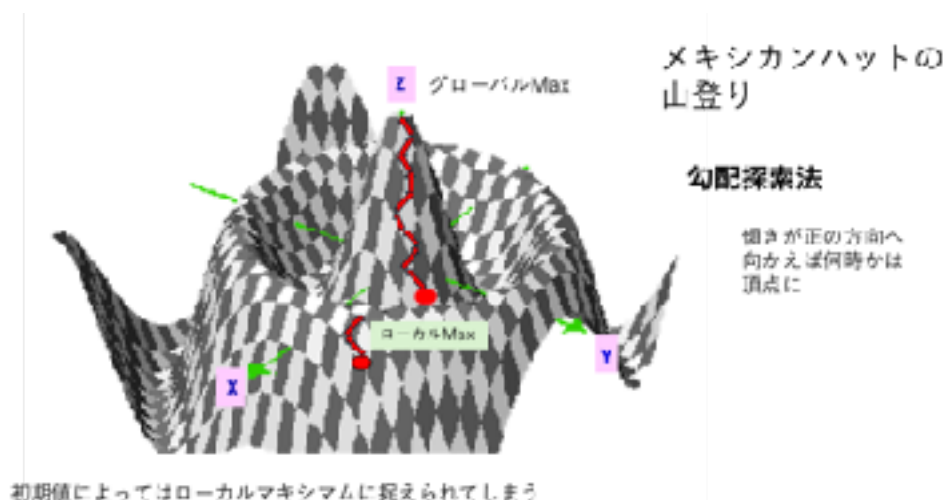
今回はClassic HansenとGA法の違いを簡単に紹介しよう。

Classic Hansen法

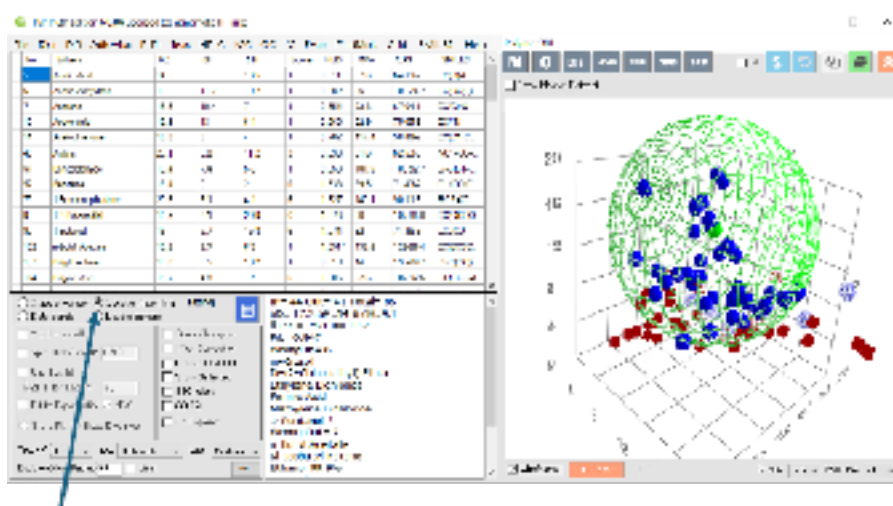


例えば例題のPolymer88J.hsdXを読み込み、Sphereを計算すると上図の結果が得られる。間違っってハンセン球の内側に入ってしまうもの、外側に出てしまうものが15溶媒ある。またSphereの半径はとても大きい。何度か計算ボタンを押すとその度に大きく変わった値が変わることがある。

これはClassic Hansen法が勾配探索法を取っている事の問題点である。



多峰性の空間探索を行う際に、ローカルマキシマムに落ち込んで、グローバルMaxに辿り着けないことがある。

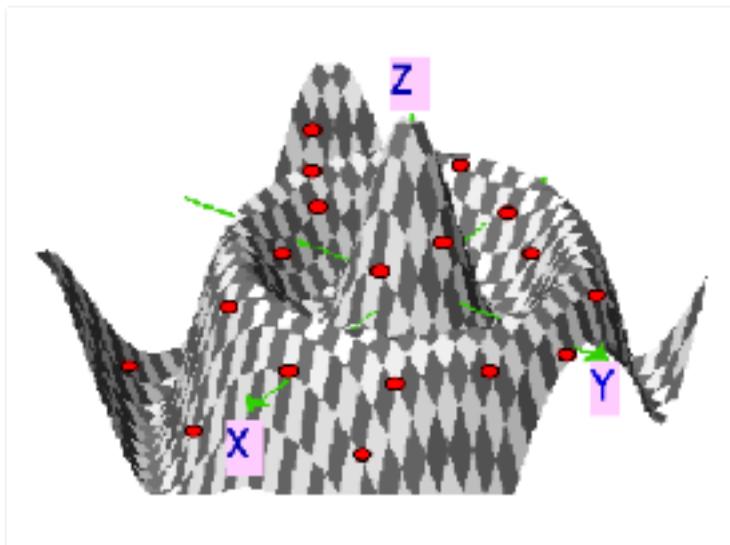


通常GA法の方がより良い解を探索できる

Genetic Algorithm(GA)法

GA法をとるとClassic Hansenより改善される。そこで、多分多くのHSPiPユーザーはGA法を使っている。（しかし、特許や論文を書く時には計算のオプションを明記すべきである）

私の作ったこの方法は空間に多くの点をばらまき、適者生存の法則で最適解を探索する。かなり昔に書いた機能なので、今回思い切って全面的に書き直した。でもやっていることは同じである。



空間に多数の初期値を置く

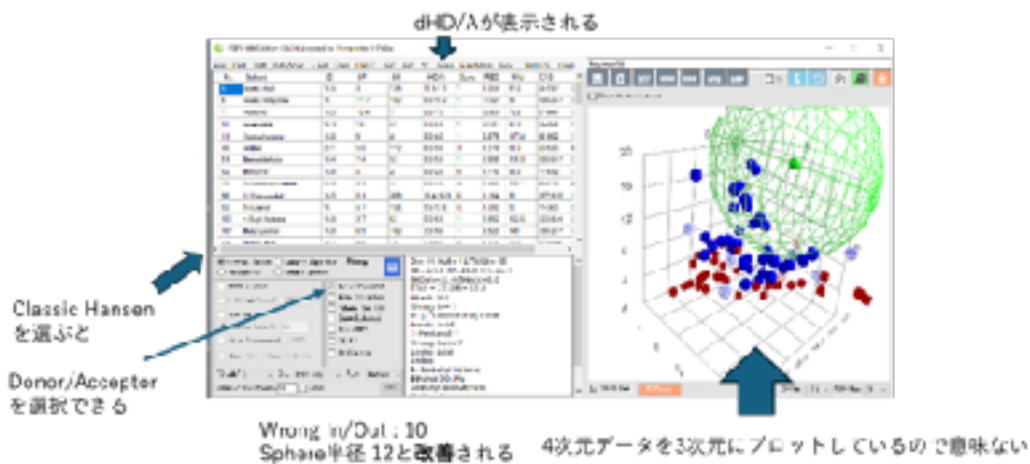
適者生存の法則

各点の評価を行う
 ルーレット選択
 突然変異
 交叉
 を繰り返す

データ数が少ない場合は答えがバタバタすることがあるが、Classic Hansenに比べれば圧倒的に良い答えを与える。

問題は水素結合項を δ_{HD} , δ_{HA} に分割した場合だ

δ_{HD} , δ_{HA} に分割した計算はClassic Hansenしか選べない。3次元プロットであれば一番



高い山がどこにあるか見当がつくかもしれない。高い山というのは、青い点が集まっているのが3次元空間上のどこか？はだいたいわかる。しかし、 $qrt(4.0*(dD1-dD2)^2+(dP1-dP2)^2+(dHacid1-dHacid2)^2+(dHbase1-dHbase2)^2)$ という式の4次元空間上で $[dD, dP, dHDon, dHAcc]=[14.3, 20.0, 11.4, 0.0]$ が一番高い山だと言われても判断のつけようがない。結果的にはWrong In/Out は10に減り、半径も12となるのでClassic Hansenよりも良さそうでもある。

ただし、大きな問題点は、 $dHDon$ の値が11.4、 $dHAcc$ が0.0になってしまうことだ。つまりこの溶質は酸性ポリマーということになる。実際にはこのポリマーはニトロセルロースであ

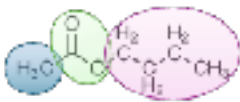



る。酸性ポリマーでは無い。このように4次元空間に勾配探索法を使うとかなりの確率でおかしなところに辿り着いてしまう。

δ HD, δ HAに分割したGA法計算はHSPiPには搭載されていない。そこでYMB24Pro4MIに搭載することにした。

ちなみにGA法で解析すると、[dD, dP, dHac, dHbs]=16.4,12.7,1.2,10.4] 半径10.5, Wrong In/Outが8となる。酸性ポリマーとはならない。

4次元データなのでSphere空間にプロットしても意味がない。そこで、GUIを必要としないMIツールとして提供することにした。

オレイン酸の溶解性

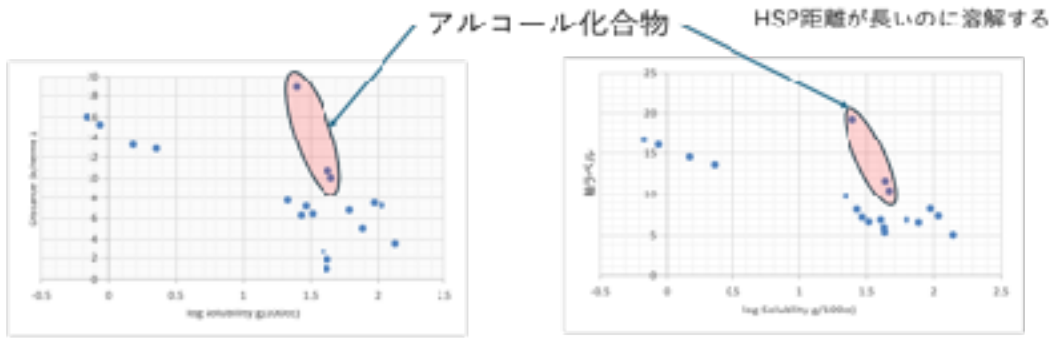
		[dD, dP, dHac]	δ H	δ Hacid	δ Hbase
中性		[15.9, 4.5, 5.9] MVol=131.9, BP=399.26K	5.9	0.1	5.9
塩基性		[15.6, 3.9, 5.8] MVol=100.9, BP=336.15K	5.8	0.41	5.8
酸性		[16.2, 3.5, 6.9] MVol=241.4, BP=585.25K	6.9	5.5	4.2
中性			6.1	3.2	4.6

まず、エステル、アミン、カルボン酸、アルコールのdHac, dHbsは上記のようになる。同じ中性のエステルとアルコールの大きな違いは、エステルはdHacがゼロであるのに対し、アルコールはdHac, dHbsともに大きな値を持つことだ。カルボン酸もともに大きな値を持つが、dHacの方が大きくなる。アミンは大きなdHbsを持ち、値を見ただけではエステルとの差が不明になる。

それでは、オレイン酸の溶解性を次の式で評価する。

$$1: \sqrt{4.0 \cdot (dD1 - dD2)^2 + (dP1 - dP2)^2 + (dH1 - dH2)^2}$$

$$2: \sqrt{4.0 \cdot (dD1 - dD2)^2 + (dP1 - dP2)^2 + (dHacid1 - dHacid2)^2 + (dHbase1 - dHbase2)^2}$$



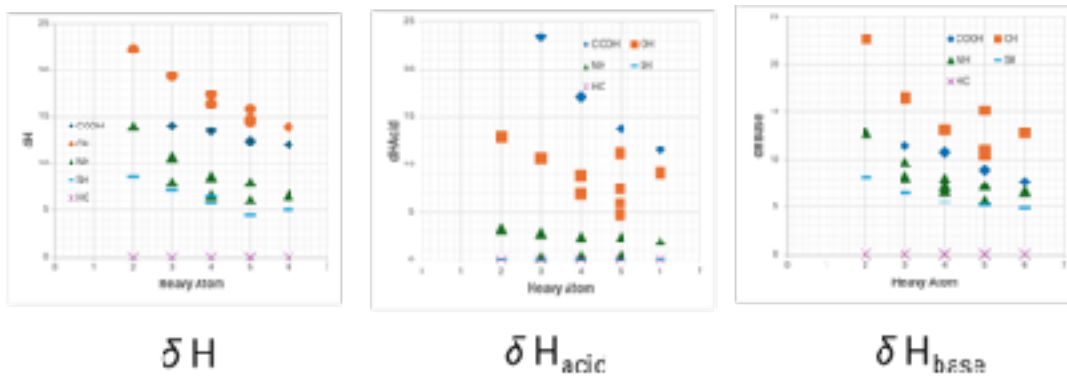
$$1: \sqrt{4.0 \cdot (dD1 - dD2)^2 + (dP1 - dP2)^2 + (dH1 - dH2)^2}$$

dHを分割しても
全体的には少し良くなったがアルコールはだめ

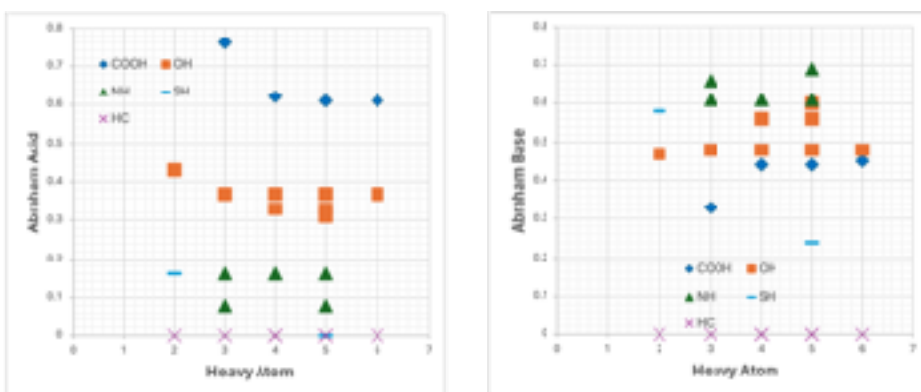
$$2: \sqrt{4.0 \cdot (dC1 - dD2)^2 + (dP1 - dP2)^2 + (dHacid1 - dHacid2)^2 + (dHbase1 - dHbase2)^2}$$

結果的には一部のアルコール化合物は、HSP距離が長いによく溶解するという変な結果になった。

これが何故起きるか？



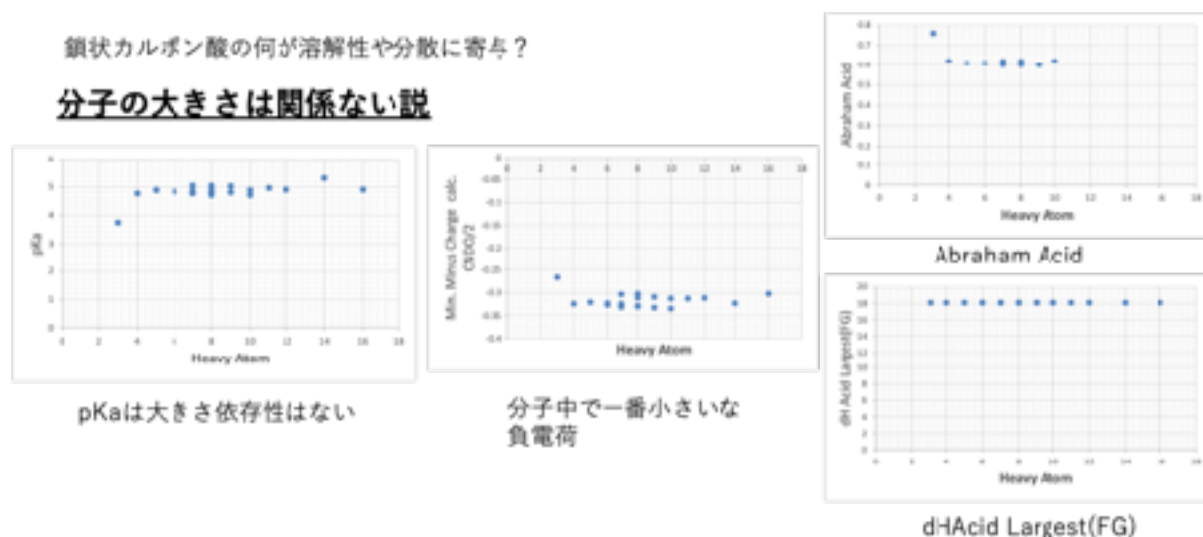
これはHSPの値は分子の大きさに依存するからである。その結果、dHac、dHbsも分子の大きさに依存する。



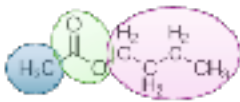



Acidになるものは、カルボン酸、アルコール、アミンの活性水素を持つものだけ。
アルコールは両性
分子の大きさには依存しない、(形状には少し依存) ➡ 分子中に存在する原子団だけで決まる

ところがAbrahamのAcid, Baseは一番小さな分子は少し影響があるが、他は分子の大きさにはほとんど影響しない。

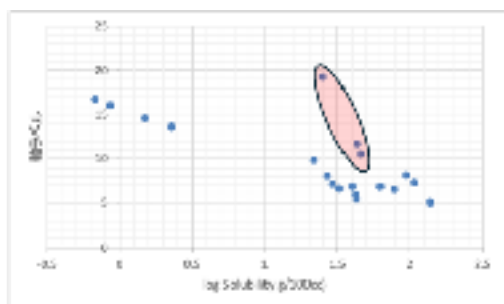
分子中にある官能基があるかないかだけでAbrahamのAcid, baseは決まっている。



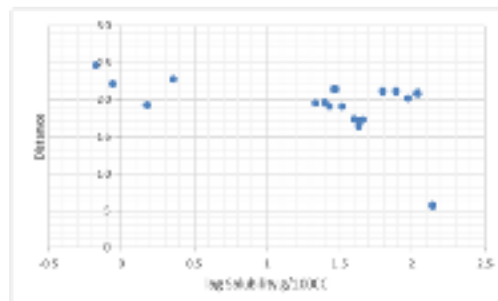
もし、溶解性や分散が分子の大きさに関係ないのだとすると、どのような指標が重要になるだろうか？ 例えばpKaの値は分子の大きさには依存性がない。MO計算で出した分子中の一番小さな負電荷も分子の大きさには依存性がない。分子を官能基に分割して、官能基ごとで一番大きなdHac、dHbsを求めれば分子の大きさには依存性がない。

			δ_H	$\delta_{+AcidLa}$	$\delta_{+BaseLa}$
中性		[15.9, 4.5, 5.9] MVol=131.9, BP=399.26K	5.9	1.1	10.7
塩基性		[15.6, 3.9, 5.8] MVol=100.9, BP=336.15K	5.8	0	14.9
酸性		[16.2, 3.5, 6.9] MVol=241.4, BP=585.25K	6.9	18.1	12.3
中性 (アルコール)		[16.1, 3.0, 6.1] MVol=323.3, BP=624K	6.1	19.3	29.2

そこで、炭化水素のアルキル基はdHacid, baseはほとんどゼロなので、dHacidLa, baseLaは上図のように求まる。これを用いてオレイン酸の溶解度を評価してみる。



$$2: \sqrt{4.0 \cdot (dD1 - dD2)^2 + (dP1 - dP2)^2 + (dHacid1 - dHacid2)^2 + (dHbase1 - dHbase2)^2}$$



$$2': \sqrt{4.0 \cdot (dD1 - dD2)^2 + (dP1 - dP2)^2 + (dHacL1 - dHacL2)^2 + (dHbaL1 - dHbaL2)^2}$$

Acidになるものは、カルボン酸、アルコール、アミンの活性水素を持つものだけ。

アルコールは両性

分子の大きさには依存しない。(形状には少し依存) ➡ 分子中に存在する原子団だけで決まる

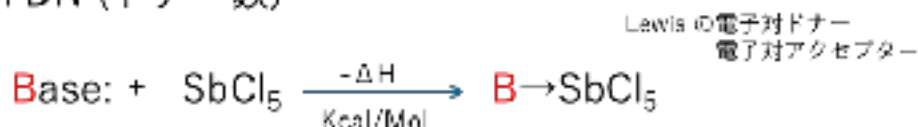
すると小さいアルコールで見られた距離が長いのによく溶解するという異常は無くなった。

Lewisのドナー、アクセプター

特に無機物の分散の時に問題となるのがLewisの定義する酸塩基だ。ここまでの話はブレンステッドの酸塩基の話で、プロトンのドナー、アクセプターだ。Lewisの定義では電子対のドナー（塩基性）、電子対のアクセプター（酸性）で、活性水素があるなしとは異なる。

GutmanはDN, ANの値を定義している。

Gutmann DN (ドナー数)



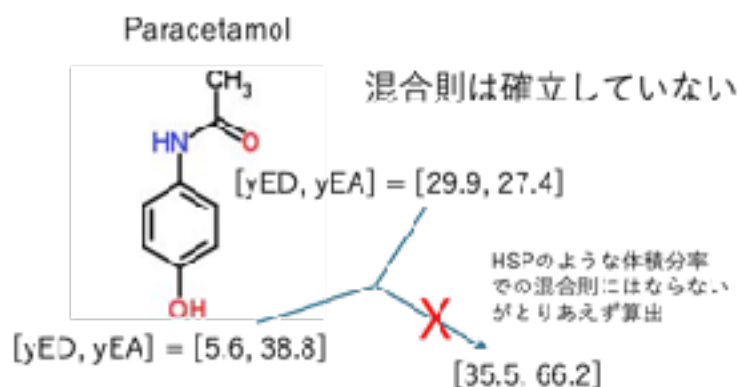
Gutmann AN (アクセプター数)



私はこのDN, ANをSmilesの構造式から求める推算式を構築し、HSP50周年記念講演会で発表している。

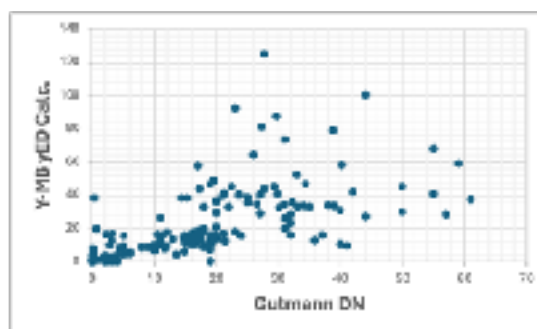
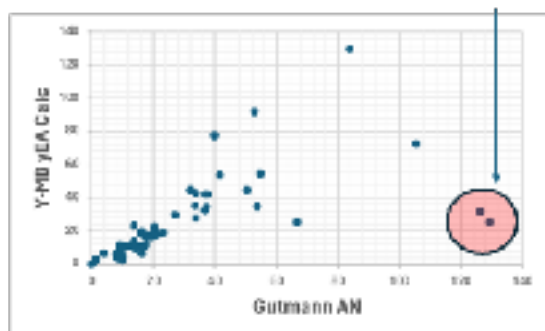
文献値のDN, ANに対して推算値のyED, yEAを計算する。(HSPiPで使われているプロトンのドナー、アクセプターと混乱しないように、YamamotoのElectron Donor, Acceptorを使っている)

複数の官能基を持つ場合の混合則は確立していないが、取り敢えずHSPと同じように体積分率で計算してみる。



すると、GutmannのDN,ANとの関係は次のようになる。大まかには良いのだが、複数の官能基を持つ化合物で上振れしてしまう。

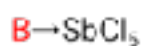
HSPと同じ混合則（体積分率）



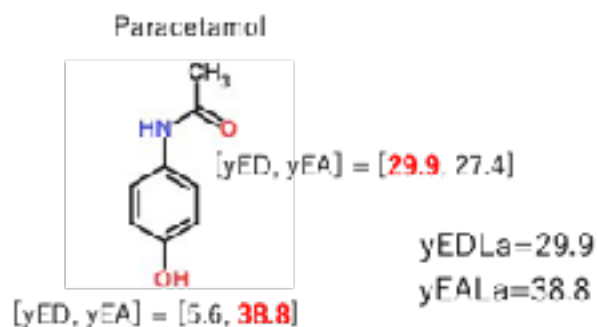
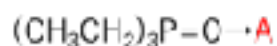
原子団にED/EAを割り振ってしまうと、分子の大きさによって計算値は上振れしてしまう

そこで、先ほどのdHAcid, dHbaseの時と同じように、分子中の官能基を調べ、一番大きな値だけを使うことを考える。

Donor Number

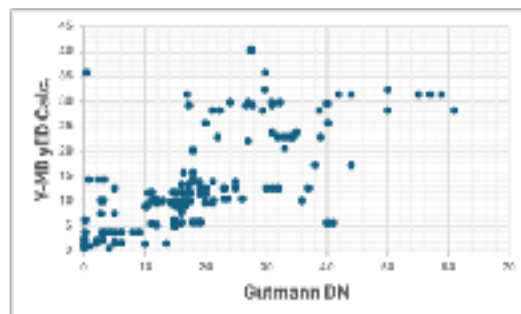
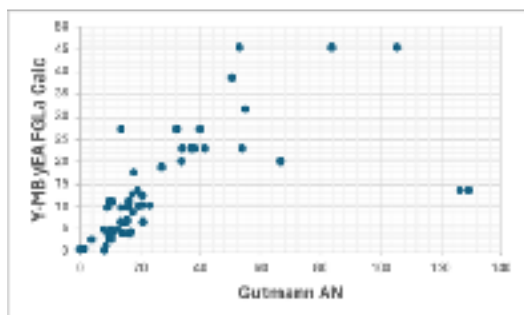


Acceptor Number



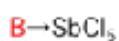
GutmannのDN,ANとの関係は次のようになる。

分子を構成する原子団で、最も大きいED, EAの計算値を出力する

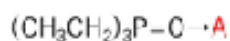


一番大きなものだけが効くと考える。

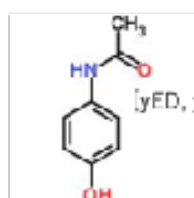
Donor Number



Acceptor Number



Paracetamol



[yED, yEA] = [29.9, 27.4]

yEDLa=29.9

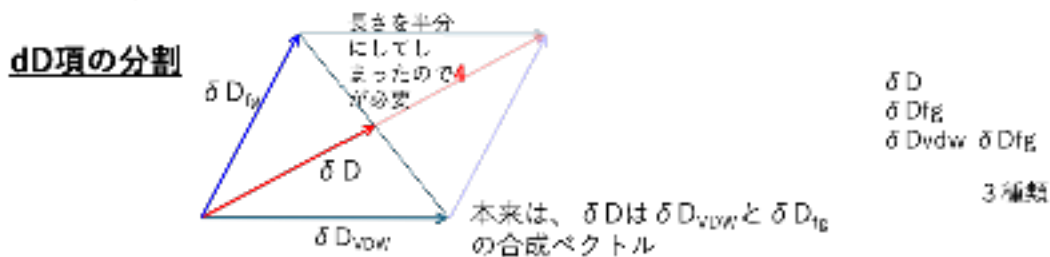
yEALa=38.8

[yED, yEA] = [5.6, 38.8]

				分子中に存在する yED, yEAの総和		
中性		[15.9, 4.5, 5.9] MVol=131.9, BP=399.26K	δ_H 5.9	yED 12.8	yEA 7.0	
塩基性		[15.6, 3.9, 5.8] MVol=100.9, BP=336.15K	5.8	44.4	6.3	
酸性		[16.2, 3.5, 6.9] MVol=241.4, BP=585.25K	6.9	14.4	26.0	
中性 (アルコール)		[16.1, 3.0, 6.1] MVol=323.3, BP=624K	6.1	11.1	12.13	

		yEDLa	yEALa
中性		10.2	4.4
塩基性		32.4	4.2
酸性		11.5	45.6
中性		12.7	23.1

2017年の拡張



水素結合項の分割(A1,B1)には

dHdo, dHac dHdoLa, dHacLa
yED, yEA yEDLa, yEALa

5*2

Abraham Acid, Base La:分子中の官能基で一番大きいもの

Euclid Type式

$$(A1-A2)^2 - (B1-B2)^2$$

Baerbower Type式

$$\text{Coeff} * (A1-A2) * (B1-B2)$$

距離がマイナスになることがある

そこで、2023年のドイツでのワークショップで発表したように、様々な距離の式を考える必要性が出てくる。

HSP距離の拡張

さまざまな式がありうる
どの式が良いのかわからない

Sphere半径の概念を捨てる

式の計算値が閾値の内側、外側を評価

Euclid Type

- $\sqrt{(dD1-dD2)^2 + (dP1-dP2)^2 + (dH1-dH2)^2}$ 1, 2 にHSPPL搭載されている
- $\sqrt{(4.0 * dDL-dD2)^2 + (dP1-dP2)^2 + (dHacid1-dHacid2)^2 + (dHbase1-dHbase2)^2}$
- $\sqrt{(dDvdw1-dDvdw2)^2 + (dDfg1-dDfg2)^2 + (dP1-dP2)^2 + (dH1-dH2)^2}$
- $\sqrt{(dDvdw1-dDvdw2)^2 + (dDfg1-dDfg2)^2 + (dP1-dP2)^2 + (dHacid1-dHacid2)^2 + (dHbase1-dHbase2)^2}$
- $\sqrt{(dDfg1-dDfg2)^2 + (dP1-dP2)^2 + (dH1-dH2)^2}$
- $\sqrt{(dDfg1-dDfg2)^2 + (dP1-dP2)^2 + (dHacid1-dHacid2)^2 + (dHbase1-dHbase2)^2}$

Baerbower Type

- $4.0 * (dD1-dD2)^2 + (dP1-dP2)^2 + \text{coeff} * (dHacid1-dHacid2)^2 + (dHbase1-dHbase2)^2$
- $(dDvdw1-dDvdw2)^2 + (dDfg1-dDfg2)^2 + (dP1-dP2)^2 + \text{coeff} * (dHacid1-dHacid2)^2 + (dHbase1-dHbase2)^2$
- $(dDfg1-dDfg2)^2 + (dP1-dP2)^2 + \text{coeff} * (dHacid1-dHacid2)^2 + (dHbase1-dHbase2)^2$
- $4.0 * (dD1-dD2)^2 + (dP1-dP2)^2 - \text{coeff} * (Y-EDL-Y-ED2) * (Y-EAL-Y-EA2)$
- $(dDvdw1-dDvdw2)^2 + (dDfg1-dDfg2)^2 + (dP1-dP2)^2 + \text{coeff} * (Y-EDL-Y-ED2) * (Y-EAL-Y-EA2)$

ドイツの時にはまとめきれていなかったが、さらに拡張した距離の式を算出する。

様々な距離の式を作り出すWebアプリ

例えばオレイン酸の溶解度のデータがある。

name	MW	Density	Temp.	Solubility	logSolubility	Score	HCode	SMILES	YamID	CAS	Score
5050 chloroform	119.378	1.482	-10	52	-1.96376732	0	156	ClC(Cl)Cl	5050	75-07-8	0
5051 acetone	58.08	0.791	-10	27	-1.43735018	0	1	CC(=O)C	5051	75-06-6	0
5052 ethanol	46.07	0.789	-10	28	-1.42854051	0	351	CCO	5052	107-03-2	0
5053 acetic acid	60.05	1.05	-10	18	-1.1384838	0	424	CC(=O)O	5053	64-19-7	0
5054 n-hexane	86.17	0.66	-10	44	-1.64786297	0	413	CCCCC	5054	110-82-3	0
5055 chloroform	119.378	1.482	-10	52	-1.96376732	1	156	ClC(Cl)Cl	5055	75-07-8	1
5056 acetone	58.08	0.791	-10	27	-1.43735018	1	1	CC(=O)C	5056	75-06-6	1
5057 ethanol	46.07	0.789	-10	28	-1.42854051	1	351	CCO	5057	107-03-2	1
5058 acetic acid	60.05	1.05	-10	18	-1.1384838	1	424	CC(=O)O	5058	64-19-7	1
5059 n-hexane	86.17	0.66	-10	44	-1.64786297	1	413	CCCCC	5059	110-82-3	1
5060 chloroform	119.378	1.482	-10	52	-1.96376732	1	156	ClC(Cl)Cl	5060	75-07-8	1

55g/100cc以上を良溶媒としよう(Score 1)。

このようなデータがあった場合、HSPiPを使わずに簡便に距離の式を作り出す方法を解説する。

New Sphereのデータ作成

CAS	Score
67-66-3	1
108-88-3	1
108-93-0	1
119-04-2	1

水の場合のCAS

CAS	
695 7732-18-5	
859 7732-18-5a	Water 1% So
527 7732-18-5b	Water Compl

Excelなどを用いて溶媒のCAS番号とScoreのテーブルを作成

HSP用のデータビルダー(Webアプリ)

New Sphereを選択

HSPiP用データ・ビルダー Js spreadsheet!

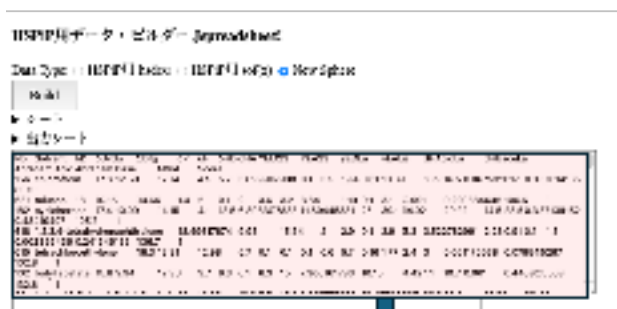
Data Type: HSPiP用 hsd(x) HSPiP用 sfd(x) New Sphere

Build クリック

シート	A	B	C	D	E	F	G	H	I	J
1	CAS	Score								
2	67-66-3	1								
3	108-88-3	1								
4	108-93-0	1								
5	119-04-2	1								

まず、溶媒のCAS番号とScoreをExcelなどで作成する。それをコピーしWebアプリ(データビルダー)に貼り付けボタンをクリックする。

入力用データを作成



CASからDBを検索

オフィシャル値があればそれを、なければY-MB24Pro4MIの推算値

ペースト

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1	No.	Solvent	dD	dDvals	dP	dH	dHD	dHA	YBDE	YBAZ	yBDA	yBAa	dHAcd	dHba	Abroa	Abroa	MVal	Score		
2	156	chloro	17.8	12.59	12.74	4.5	5.7	5.6999	0.1	3.5	19.4	0.61	10.39	18.9	16.5	0.0875	0.0590	80.5	1	

CAS番号からI.2Kデータベースを検索し、オフィシャル値があればそれを使い、オフィシャル値が無ければYMB24Proの推算値を埋めたテーブルを作成する。3種類のdDとdHの分割(5タイプ)が含まれる。



各計算式でのWrong In/Outの最小値となるターゲットのHSPを求める



それをSphere計算用Webアプリに投げ込むと各種計算式でWrong In/outが最小になる溶質のHSP²と半径を探索する。

Euclidタイプ(差分の2乗, dDの前には4の係数)の18種類の式が評価される。

20240708 ver Euclid Distance sqrt(coeff*(HSP1-HSP2) ² . . .)						
SSEc1-Wrong In/	dD	dP	dH	Radius		

2	17.17	2.61	3.21	4.04		
SSEc2- Wrong In/	dDfg	dP	dH	Radius		
2	13.33	2.00	3.22	3.58		
SSEc3- Wrong In/	dD	dP	dHac	dHba	Radius	
2	17.20	2.72	2.31	2.53	4.70	
SSEc4- Wrong In/	dD	dP	dHacLa	dHbaLa	Radius	
3	17.25	1.63	3.52	6.11	5.88	
SSEc5- Wrong In/	dDfg	dP	dHac	dHba	Radius	
2	12.82	2.30	2.90	2.49	4.29	
SSEc6- Wrong In/	dDfg	dP	dHacLa	dHbaLa	Radius	
3	13.79	1.55	3.53	6.08	5.70	
SSEc7- Wrong In/	dD	dP	yED	yEA	Radius	
4	16.41	1.65	5.48	7.21	7.36	
SSEc8- Wrong In/	dD	dP	yEDLa	yEALa	Radius	
2	17.97	1.68	4.66	5.99	6.62	
SSEc9- Wrong In/	dDfg	dP,	yED	YEA	Radius	
4	13.35	1.70	5.48	7.31	7.41	
SSEc10- Wrong In/	dDfg	dP	yEDLa	yEALa	Radius	
2	12.97	2.07	2.67	5.96	5.63	
SSEc11- Wrong In/	dD	dP	AbraAcid	AbraBase	Radius	
2	17.28	3.15	0.06	0.19	3.64	
SSEc12- Wrong In/	dDfg	dP	AbraAcid	AbraBase	Radius	
2	12.80	3.18	0.10	0.34	3.16	
SSEc13- Wrong In/	dDvdw	dDfg	dP	dH	Radius	
2	11.32	13.30	2.07	3.13	3.81	
SSEc14- Wrong In/	dDvdw	dDfg	dP	dHac	dHba	Radius

2	10.76	12.83	1.90	3.47	2.75	4.83
SSEc15- Wrong In/	dDvdw	dDfg	dP	dHacLa	dHbaLa	Radius
3	10.24	13.79	1.58	3.72	5.84	5.88
SSEc16- Wrong In/	dDvdw	dDfg	dP	yED	yEA	Radius
4	9.95	13.58	2.05	5.36	7.60	7.83
SSEc17- Wrong In/	dDvdw	dDfg	dP	yEDLa	yEALa	Radius
3	11.80	13.95	2.04	3.26	4.01	4.68
SSEc18- Wrong In/	dDvdw	dDfg	dP	Abra Acid	Abra Base	Radius
2	11.26	12.69	2.97	0.13	0.26	3.46

多くの式でWrong In/Outは2であった。

この、Euclidタイプの式では、HSP距離が最小になるには、溶媒と溶質のHSPが完全に一致した時になる。そこで酸-塩基相互作用は考慮されない。

次にBeerbowerタイプの式が出力される。

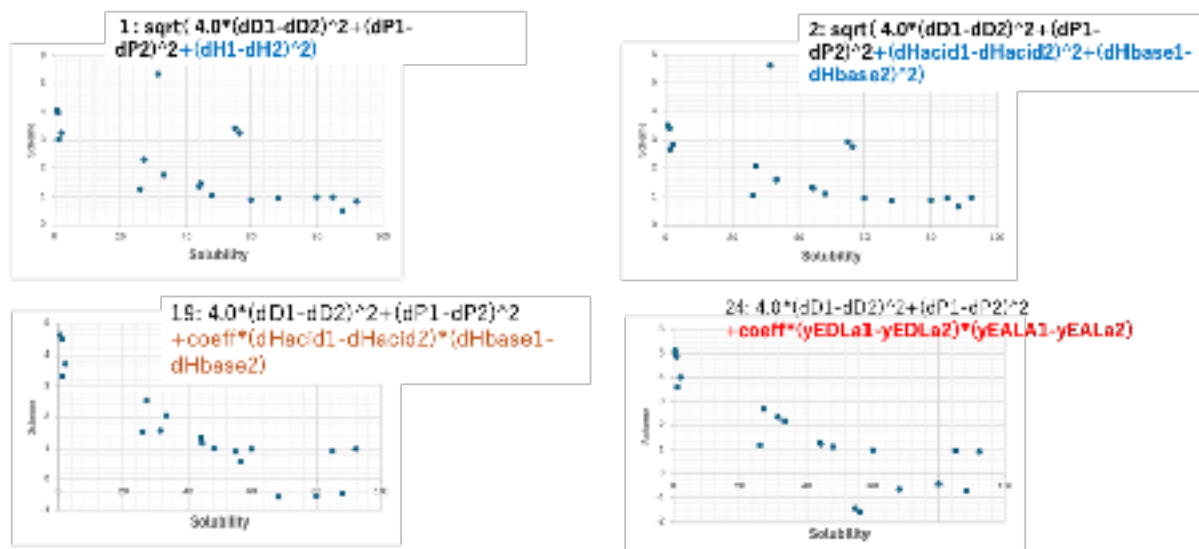
Beerbowerタイプでは距離の式は、 $\text{coeff} \cdot (A1-A2) \cdot (B1-B2)$ で評価される。例えばこの式がマイナスになるには、溶媒のA1, B1のどちらかが溶質のA2, B2より大きい必要がある。例えばカルボン酸がダイマーを作っているとしよう。水素結合なので自分自身のA2とB2で相互作用している。そこにA1がA2よりも大きい溶媒が来ると、B2は自分のA2と相互作用するより、A1と相互作用する方が有利になる。弱酸の塩に強酸を加えると交換するようなものだ。

系によっては式全体の評価がマイナスになるのでルートは取らない。式が閾値より小さいと考えるので正確にはRadiusではないしRadiusがマイナスのこともある。

Beerbower Distance $\sqrt{\dots \text{coeff} \cdot (Acd1-$						
SSBeB1 9-Wrong	dD	dP	dHac	dHba	Coeff	Radius
0	17.40	1.82	3.42	5.91	-0.44	3.73
SSBeB2 0-Wrong	dD	dP	dHacLa	dHbaLa	Coeff	Radius
1	18.45	2.60	6.43	10.31	-0.46	-1.61
SSBeB2 1-Wrong	dDfg	dP	dHac	dHba	Coeff	Radius
1	13.14	2.78	3.14	5.00	-0.19	2.36

SSBeB2 2-Wrong	dDfg	dP	dHacLa	dHbaLa	Coeff	Radius	
1	10.62	2.42	10.25	11.20	-0.39	-1.44	
SSBeB2 3-Wrong	dD	dP	yED	yEA	Coeff	Radius	
1	17.46	3.77	10.49	20.53	-0.16	-0.09	
SSBeB2 4-Wrong	dD	dP	yEDLa	yEALa	Coeff	Radius	
0	17.11	3.17	6.33	8.20	-0.39	3.10	
SSBeB2 5-Wrong	dDfg	dP,	yED	YEA	Coeff	Radius	
1	13.84	2.94	7.40	19.94	-0.09	2.16	
SSBeB2 6-Wrong	dDfg	dP	yEDLa	yEALa	Coeff	Radius	
1	13.19	3.41	7.33	10.47	-0.39	1.36	
SSBeB2 7-Wrong	dD	dP	AbraAci d	AbraBa se	Coeff	Radius	
2	17.19	3.58	0.10	0.24	0.13	3.76	
SSBeB2 8-Wrong	dDfg	dP	AbraAci d	AbraBa se	Coeff	Radius	
2	13.76	2.85	0.13	0.23	-2.43	3.19	
SSBeB2 9-Wrong	dDvdw	dDfg	dP	dHac	dHba	Coeff	Radius
0	13.43	12.88	2.18	3.86	3.96	-0.29	4.11
SSBeB3 0-Wrong	dDvdw	dDfg	dP	dHacLa	dHbaLa	Coeff	Radius
1	13.35	15.43	2.79	7.95	10.15	-0.44	-2.14
SSBeB3 1-Wrong	dDvdw	dDfg	dP	yED	yEA	Coeff	Radius
1	12.04	14.27	2.90	6.31	21.60	-0.21	2.03
SSBeB3 2-Wrong	dDvdw	dDfg	dP	yEDLa	yEALa	Coeff	Radius
1	12.97	14.15	2.67	5.41	10.56	-0.48	2.26
SSBeB3 3-Wrong	dDvdw	dDfg	dP	Abra Acid	Abra Base	Coeff	Radius
2	11.57	12.50	3.00	0.06	0.35	1.90	3.51

SSBeB19、SSBeB24、SSBeB29でWrong In/Outがゼロになる。

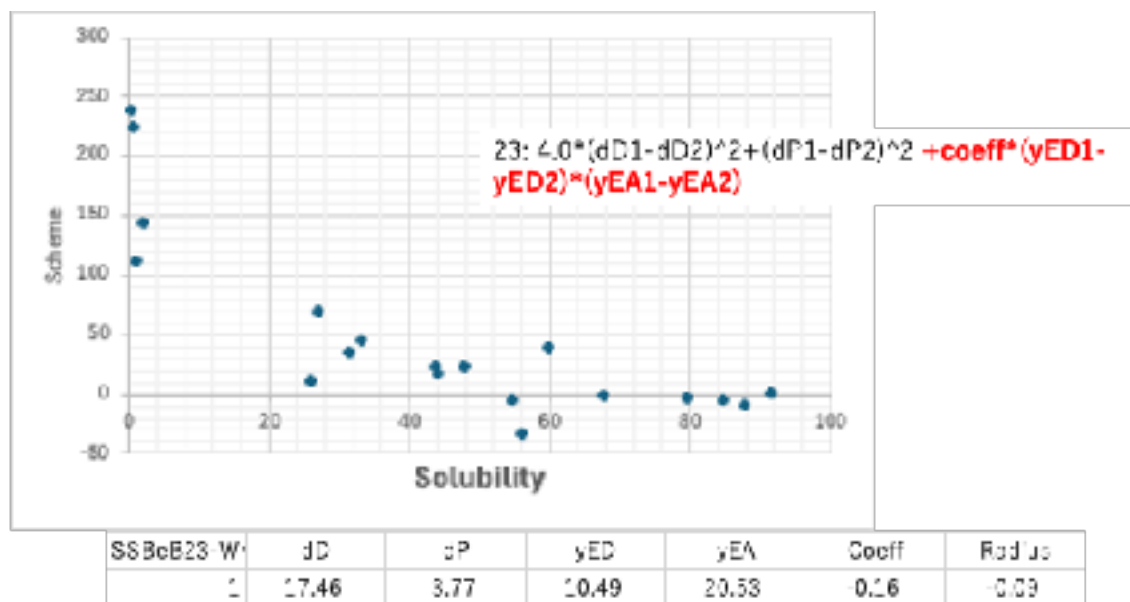


式1、2はHSPiPに搭載されている計算式をGA法で求めたものになる。やはり一部の小さなアルコールで変なことが起こる。

Beerbowerタイプの式では、かなり理解しやすい結果となる。

ただし、我々はオレイン酸は酸性であることを知っている。つまり、dHacidかyEAが大きなもののが答えのはずである。

そこで、Wrong In/outは1ではあるが次の答えが一番納得がいく。



Wrong In/outは1

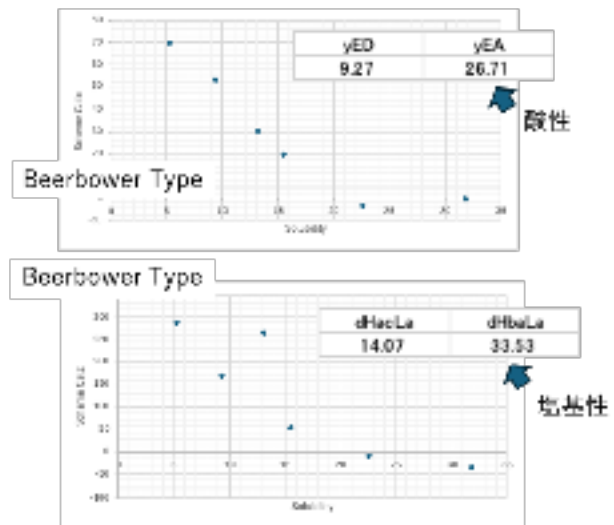
この化合物は酸性

これは酸であることを知っているのだから選べるのであってそうで無ければ選べない。いつまで経っても、AIには選べず、人間の研究者だけが33種類の式から判断していくことになる。

データは少ないがオクタデシルアミンの溶解度はどの式もwrong In/Outはゼロになる。

データ数は少ないがオクタデシルアミン

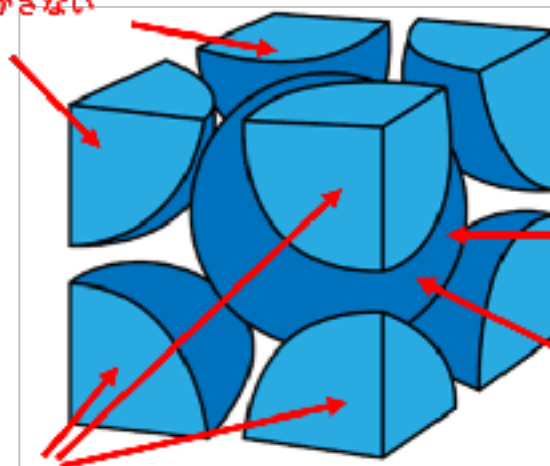
YeastID	name	PH	Observed	Temp.	Solubility
5080	Acetate de Rie	41.023	0.736	20	
5127	Zarblum	18.016	0.79	20	
5127	no propyl alcohol	80.000	0.40	20	
5128	amphibia	136.167	0.69	20	
5129	al denon am	115.258	1.106	20	
5050	al denon am	115.278	1.426	20	31.9
5172	sulfuric acid	85.109	0.500	20	5.5
5280	barbiturate	118.19	0.60	20	
5042	no propyl alcohol	151.013	1.534	20	
5140	cocoholone	84.161	0.770	20	13.2
5171	Li-donora	85.1	1.020	20	
5182	Allyl alcohol	71.123	0.112	20	
5080	al denon am	10.000	1.70	20	
5171	r (2) alcohol	13.58	1.578	20	
5080	al denon am	61.041	1.120	20	
5174	r (2) alcohol	121.112	1.130	20	
5470	alkali	84.000	1.100	20	
5183	isotonic	71.123	0.51	20	29.7
5182	Allyl alcohol	71.127	0.000	20	
5171	r (2) alcohol	81.117	0.000	20	0.1
5470	alkali	75.111	0.070	20	
5080	al denon am	20.000	0.491	20	15.0



溶質が酸であるとしても(yEAが大きい)、塩基である(dHbaseが大きい)としても解釈できてしまう。

Sphereの理論 Again

ポリマーを溶かさない



粒子の遅い沈降性

(拡張された)
HSPが近いものが
同じような性質を持ち
Hansen空間中で集まる

ポリマーの**良い**溶解性
が集まることもある

粒子の**早い**沈降性
が集まることもある

一般にとっても誤解されているのが、Sphere, ハンセンの溶解球の理論だ。
 溶解球の理論はHSPが近いものが**同じような性質**を持ちHansen空間中で集まるとだけ言っている。この同じような性質というのが、ポリマーを良く溶解するだった場合は溶解球の外側には溶かさないという性質の溶媒が配置される。

LumisizerやパルスNMRの時の**相対沈降時間**や緩和時間はどうか？

粒子の**早い沈降性**の溶媒が集まっていることがあった。

すると遅い沈降性の溶媒は溶解球の外側にくる。

その時に、良いと悪いの人間にとっての評価はどうか？

化粧品のように沈殿して欲しく無ければ溶解球の外側にある溶媒を良溶媒と呼ぶだろう。粉体を沈殿させてケーキを作りたいところは、早い沈降性のものを良溶媒と呼ぶだろう。Scoreによってみるものが全く異なる。



CBの表面理解に欠かせない

緩和時間が長い方がHSP距離短い

$$[dD, dP, dH] = [17.4, 10.5, 6.3]$$

← 極性溶媒ほどHSP距離が短くなり緩和時間が長くなる

$$[dD, dP, dHAcid, dHBase] = [17.0, 10.3, 0.2, 8.2]$$

緩和時間が短い方がHSP距離短い

クラッシュック [dD, dP, dH] では説明できない

$$[dD, dP, dHAcid, dHBase] = [15.9, 4.0, 0.182, 0.596]$$

$$[yD, yP, yED, yEA] = [21.1, 5.0, 16.4, 17.2]$$



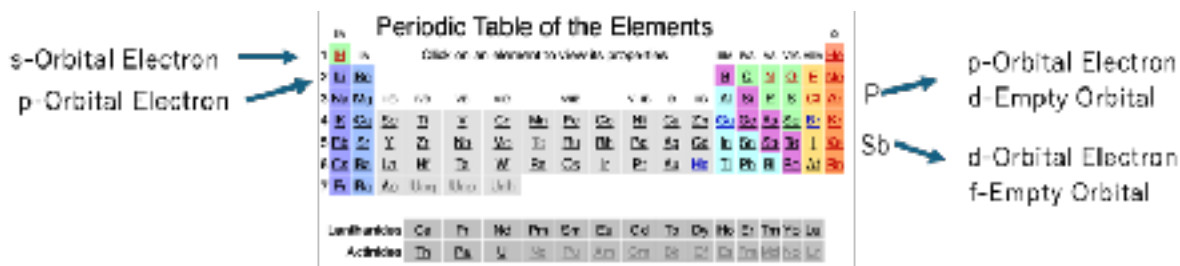
Butanol	0.42
Monoethanolamine	0.37
Hexane	0.33
Dodecane	0.2
Ethylene Glycol	0.19
n-Pentane	0.04

Acid-Base, yED-yEAの交換作用で距離を縮めなければ理解できない

そのような場合にはScoreの値を反転させて、距離の式を作ってみれば何が起きているかはっきりする。

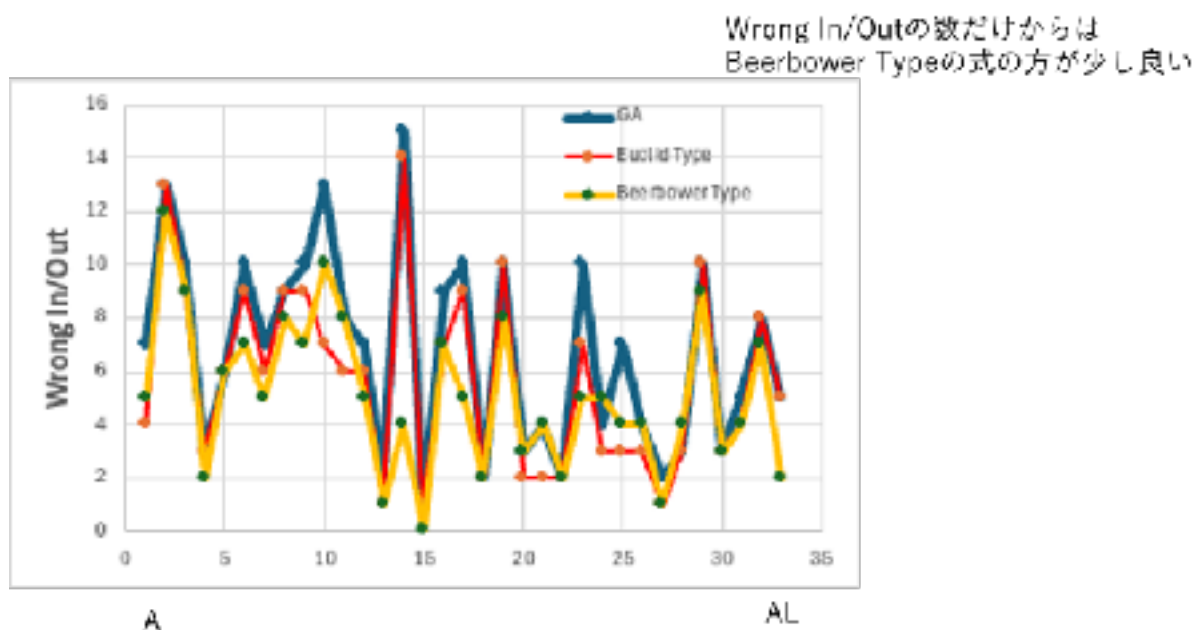
現実にはパルスNMRで緩和時間が短いほうを良溶媒と呼んだり、短い方を良溶媒と呼んだり混乱している。

こうした無機物微粒子は周期律表でFより後ろの原子を含むことが多い。



このd軌道の空軌道やf軌道の空軌道が方向性を持っていて分子全体の平均として働くのではない。方向性を持つ溶解、分散になる時にyED, yEAは大きく関与する。

この新しい距離の式をHSPiPの例題、Polymer88A-AIに適用してみた。



この例題は溶媒88種類を使って溶解度試験をしたものだ。

Beerbowerタイプの式の方が少し良い。

その中でも実際に性能の高かった分割法は以下のものであった。

こうした評価が進めば、33種類の式も淘汰されていくかもしれない。

Euclid TypeでClassicな分割でもそこそこ性能が出る

Euclid Type

- 1: $\text{sqrt}(4.0*(dD1-dD2)^2+(dP1-dP2)^2+(dH1-dH2)^2)$ 36%
- 2: $\text{sqrt}(4.0*(dD1-dD2)^2+(dP1-dP2)^2+(dHacid1-dHacid2)^2+(dHbase1-dHbase2)^2)$ 30%
- 3: $\text{sqrt}((dDvdw1-dDvdw2)^2+(dDfg1-dDfg2)^2+(dP1-dP2)^2+(dH1-dH2)^2)$ 18%

1: $\text{sqrt}((dDvdw1-dDvdw2)^2+(dDfg1-dDfg2)^2+(dP1-dP2)^2+(dHacid1-dHacid2)^2+(dHbase1-dHbase2)^2)$
 2: $\text{sqrt}((dDfg1-dDfg2)^2+(dP1-dP2)^2+(dH1-dH2)^2)$
 3: $\text{sqrt}((dDfg1-dDfg2)^2+(dP1-dP2)^2+(dHacid1-dHacid2)^2+(dHbase1-dHbase2)^2)$

Beerbower Type

- 7: $4.0*(dD1-dD2)^2+(dP1-dP2)^2+\text{coeff}*(dHacid1-dHacid2)*(dHbase1-dHbase2)$ 39%
- 8: $((dDvdw1-dDvdw2)^2+(dDfg1-dDfg2)^2+(dP1-dP2)^2+\text{coeff}*(dHacid1-dHacid2)*(dHbase1-dHbase2))$
- 9: $((dDfg1-dDfg2)^2+(dP1-dP2)^2+\text{coeff}*(dHacid1-dHacid2)*(dHbase1-dHbase2))$
- 10: $4.0*(dD1-dD2)^2+(dP1-dP2)^2+\text{coeff}*(Y-ED1-Y-ED2)*(Y-EA1-Y-EA2)$ 42%
- 11: $((dDvdw1-dDvdw2)^2+(dDfg1-dDfg2)^2+(dP1-dP2)^2+\text{coeff}*(Y-ED1-Y-ED2)*(Y-EA1-Y-EA2))$

Beerbower Typeではポリマー自体の極性がある場合、効果的

しかし、さらに、次の式が試されようとしている。

$$\text{sqrt}(4*(dD1-dD2)^2+(dP1-dP2)^2+(dH1-dH2)^2) + \text{coeff}*(A1-A2)*(B1-B2)$$

元々、HSP自体は良い性能を持っている、例外になるドナーアクセプター相互作用は式の外に出す方が合理的なので、片っ端から式を作って評価してしまう。また、Double Sphereや実データを使ったSphere法の開発も並行している。

MIユーザーは確固としたセオリーがあるかどうかはあまり気にしない。データ駆動型で出てきた答えが合理的であれば良い。その判断基準がクラシックなdD, dP, dHから大幅に改善され始めた。

まだまだ開発中でベータ・バージョンであるが一部のMIユーザーは利用を始めている。今月のpirika newsは使い方マニュアルのようなものになってしまった。例題のExcelのファイルとともに利用してほしい。