

【Hydrogen Bond Propensityの使い方】

2020年5月18日
CSD2020使用

目的: 水素結合のPairingを予測する。

自分の分子がよくある水素結合様式をしているかどうかを確認する。

手順: Target分子を元に類似のCSDデータを抽出する。

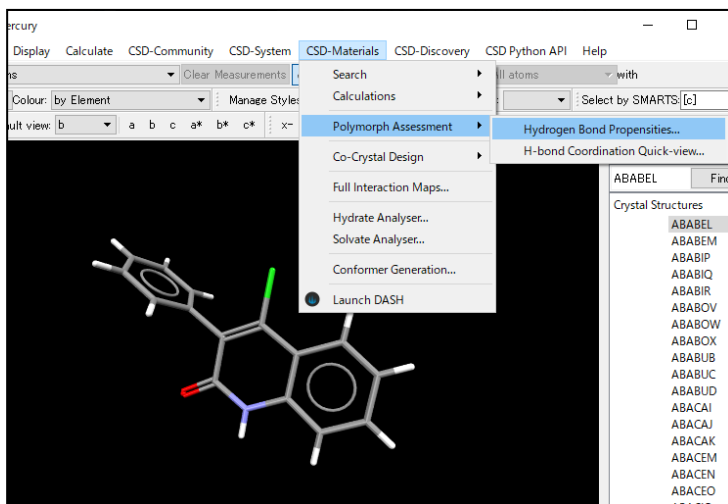
Donor, Acceptor部位を判別する。

水素結合の有無を判別する。

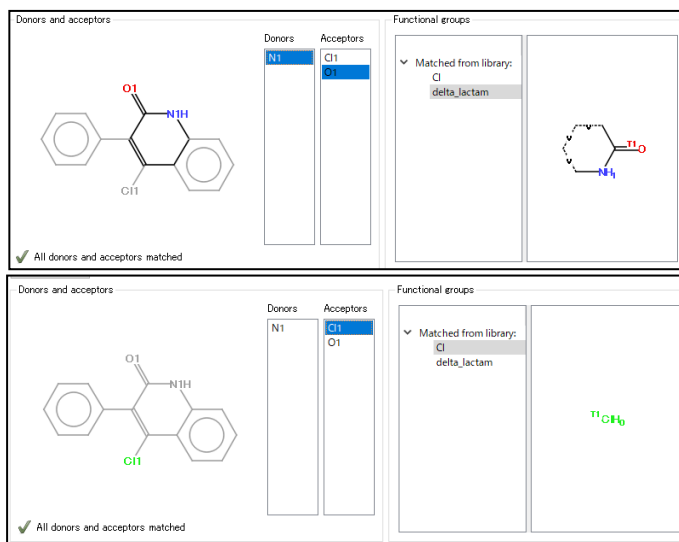
環境(Chemical group, Competition, Stericなどのfactor)を考慮の上, モデルを構築する。
モデルを元に水素結合Pairingを予測し, チャートに示す。

【例 1】 まずは, 水素結合部位が少ない分子で練習

1. CSD-Materials→Polymorph Assessment→Hydrogen Bond Propensitiesより起動。
(ここでは, ABABELを使用)

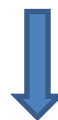


2. 水素結合に関与するDonor, Acceptorが自動認識される。(ここでは2Dとして認識)



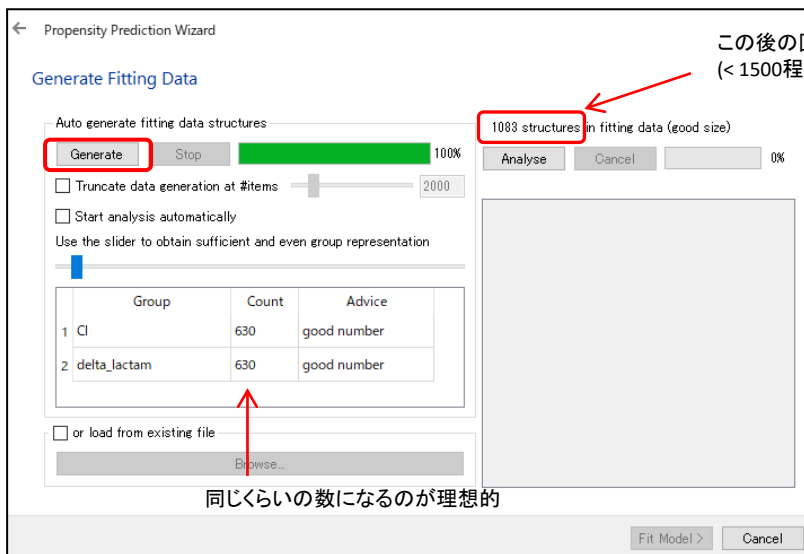
N1: Donor
O1: Acceptor
Cl1: Acceptor

統計処理をするにあたり
十分な数ヒットする骨格。
ここではラクタム骨格

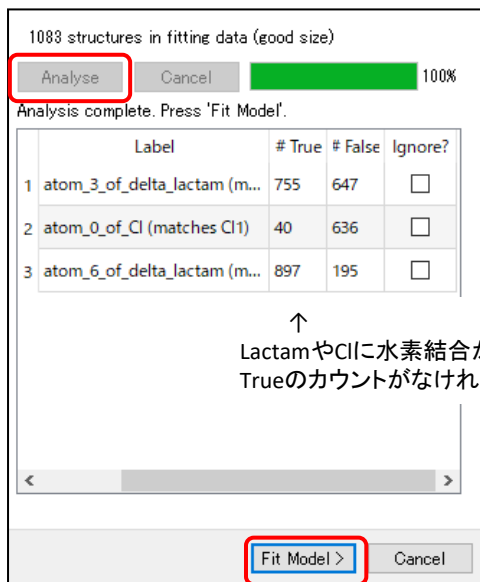


これらのfragmentを持つ
CSDのデータを抽出する。

3. Generateを押すとCSD中に含まれる”Group”(水素結合する置換基)を持つstructureを検索し、カウントが始まる。完了すると、十分な数あったかどうか(good number)判定される。またfitting dataに使うのに適した数に自動で抽出される(ここでは1083件)。



4. Analyseを押すと抽出したCSDデータに対し、実際に水素結合していたかどうか分析し、True or Falseに水素結合のカウントが表示される。その後、Fit Modelをクリック。



注: Perfect prediction. 件数が< 5の場合、赤字で警告される。その場合、Ignoreにする。

【理由について気になる方は】

<https://stats.idre.ucla.edu/other/mult-pkg/faq/general/faqwhat-is-complete-or-quasi-complete-separation-in-logisticprobit-regression-and-how-do-we-deal-with-them/>

5. Hydrogen Bond Propensity Scoreに使用するLogit modelが構築される。 問題なければ、Accept & Calculateへ。

Propensity Prediction Wizard

Model Fitting

Use this page to **fit**, **assess** and **refine** a hydrogen bond logit model.

Model Coefficient Statistics

logit_model_1 Coefficients:							
Coefficients:	Estimate	Std. Error	z value	Pr(> z)	Significance code	Lower Bound	Upper Bound
(Intercept)	-2.477	0.288	-8.599	8.03441e-18	***	-3.053	-1.922
Donorother	1.328	0.104	12.710	5.18449e-37	***	1.125	1.534
Acceptoratom_6_of_delta_lactam	5.843	0.229	25.509	1.55529e-143	***	5.409	6.309
Acceptorother	4.434	0.207	21.405	1.19985e-101	***	4.045	4.860
Competition	0.050	0.015	3.360	0.000778649	***	0.021	0.080
Donor_steric_density	-0.010	0.003	-3.965	7.3549e-5	***	-0.015	-0.005
Acceptor_steric_density	-0.034	0.003	-11.320	1.04462e-29	***	-0.040	-0.028
Donor_aromaticity	-0.395	0.312	-1.074	0.282948		-0.046	0.278
Acceptor_aromaticity	-0.648	0.302	-2.149	0.0316392	*	-1.242	-0.060
Donoratom_3_of_delta_lactam	0.000	N/A	N/A	N/A	N/A	N/A	N/A
Acceptoratom_0_of_C	0.000	N/A	N/A	N/A	N/A	N/A	N/A

Area under ROC curve = 0.859385 (excellent discrimination)

Area under ROC curveは、1に近いほどよい。

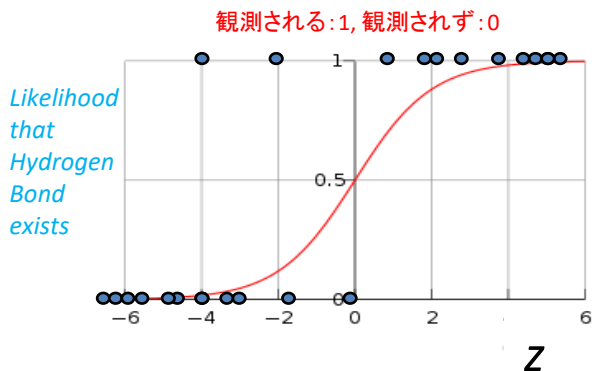
Accept & Calculate > Cancel

ここではfragmentの2D情報を元にモデルを構築。
→分子の3D構造に依存せず

Signification codeに「*」なしの場合、モデルに寄与していないのでRefine Modelのボタンから削除してもok。

正: 水素結合に寄与
負: 水素結合することにマイナスに働く

【参考】 Logistic regressionについて



$$\text{Propensity (bond exists)} = \frac{1}{1 + e^{-z}}$$

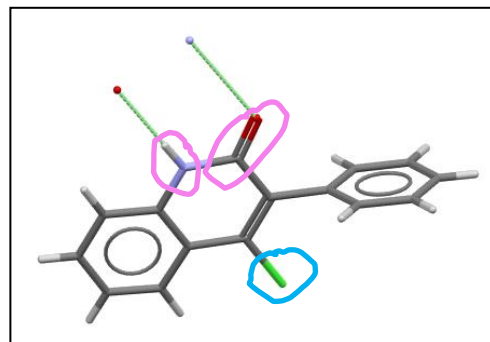
1に近いほど、よく観測される

$$Z = \theta_0 + \theta_1 \cdot \text{Donor Group1} + \theta_2 \cdot \text{Donor Group2} + \dots + \theta_3 \cdot \text{Acceptor Group 3} + \theta_4 \cdot \text{Acceptor Group 2} + \dots + \theta_5 \cdot \text{Competition} + \theta_6 \cdot \text{Donor Steric Density} + \theta_7 \cdot \text{Acceptor Steric Density} + \theta_8 \cdot \text{Donor Aromaticity} + \theta_9 \cdot \text{Acceptor Aromaticity}$$

Parameters which describe the environment of the functional group and allow customisation of calculation to target molecule. (ターゲット分子毎にパラメーターが異なる)

★HB Propensity model: つど計算される。2DIに関わるfactorのみ。

詳細: P.T.A. Galek *et al.*, *CrystEngComm*, **11**, 2634-2639, (2009).



6. 先ほどのLogit modelにもとづき, Hydrogen Bond Propensity (x軸) が計算され, Co-ordination score (y軸)と共に表示される.

Propensity scores
Hydrogen-bond propensities for ABABEL calculated using 'logit_model_1'

	Donor	Acceptor	Propensity	ABABEL
1	INTER-molecular			
2	N1 of delta_lactam	O1 of delta_lactam	0.65	observed
3	N1 of delta_lactam	Cl1 of Cl	0.00	

Left-clicking table items visualizes H-bond groups/contacts and interacts with chart Atom Highlighting...

【Propensity Score】

HB Propensity Scoreは, (ざっくり言って)水素結合していたかの頻度(確率). 水素結合能力.

今回, 読み込んだ分子の場合, INTER-molecular相互作用としては, N1とO1で0.65のところ, 実際にもobserved. N1とCl1は0.00で, 観測されず.

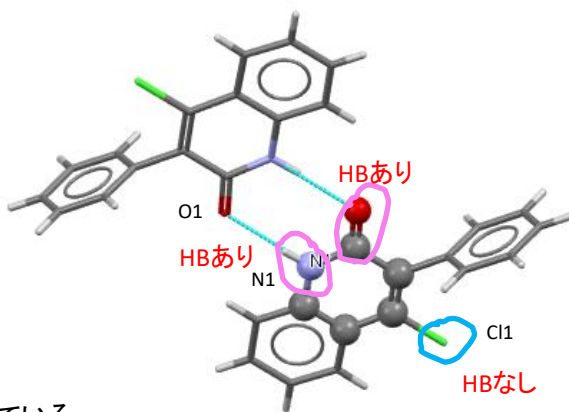
Co-ordination scores
(To refresh table: left-click chart point)

Atom (D/A)	= 0	= 1	= 2	= 3
1 N1 of cyclic_am...	0.009	0.960	0.031	0.000
2 Cl1 of Cl (a)	0.973	0.026	0.001	0.000
3 O1 of cyclic_am...	0.138	0.743	0.115	0.005

【Co-ordination Score】

Co-ordination Scoreとは, 水素結合するAtomに対し, 実際にどのように水素結合していたかどうかのスコア. HBネットワーク様式を予測するための値. =0, =1, =2, =3は, 1つのAtomに対しての水素結合の数. N1とO1はHBあり. =1のScoreが一番高い. Cl1は, HBなし. Scoreも=0が一番高い.

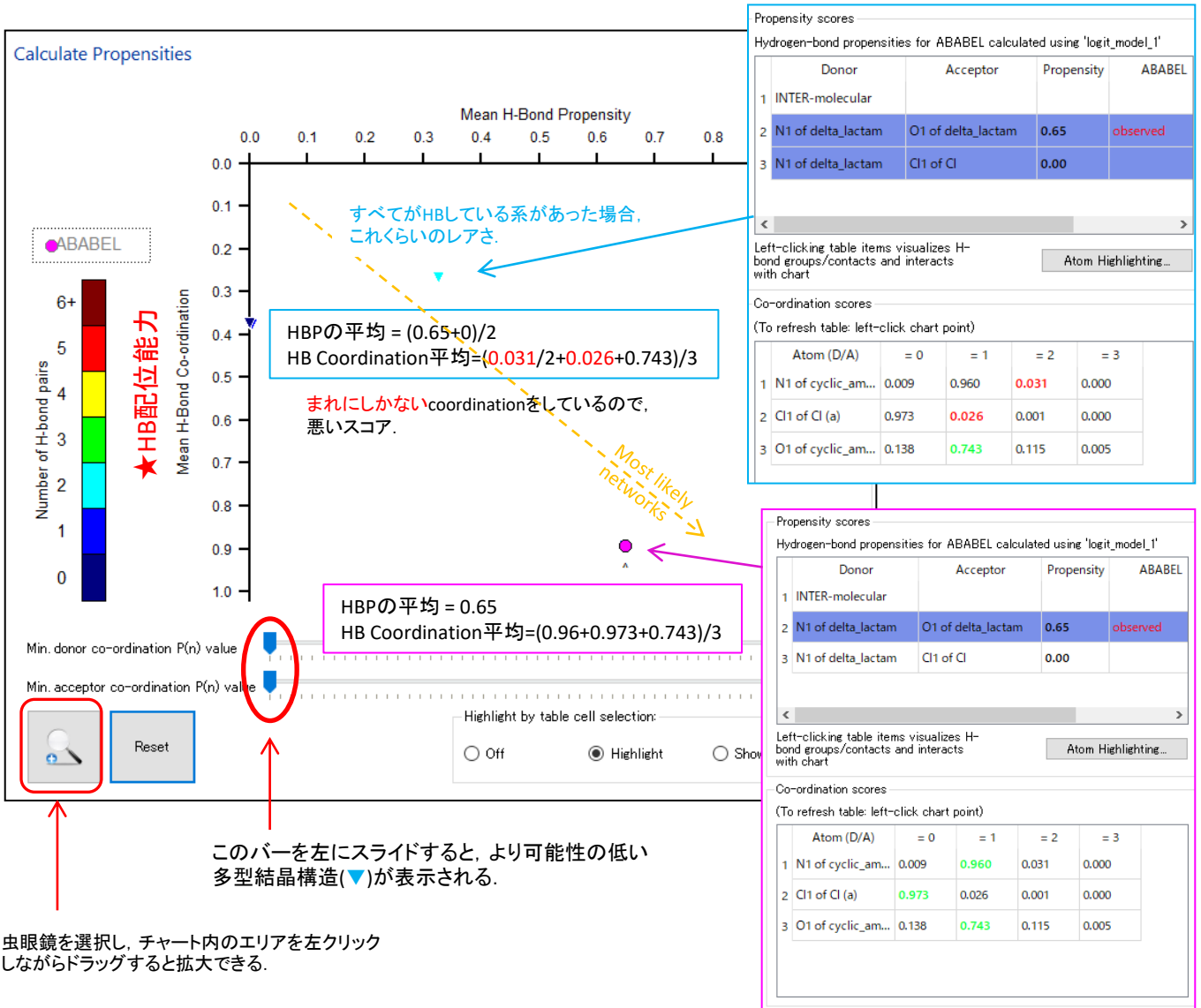
↑
ハイライトされているのは, 読み込んだ結晶構造でのHB様式. 緑は, 最適な様式. (レアな様式が観測された場合, 赤)



Co-ordination Scoreとは, Propensity同様, HBするfragmentを自動認識し, 事前に計算してあるモデルを基にScoreが計算される. HBP Scoreと異なる点は, 分子配座(3D)の寄与あり. →配座により水素結合様式が変わる可能性も加味されている. ここでは説明を省略するが, 詳細は, 下記文献を参照.

【詳細】 P.T.A. Galek *et al.*, *CrystEngComm*, **11**, 2634-2639, (2009).

7. チャートは、HB Propensity(平均) vs. Co-ordination score (平均)をプロットしたもの。
 →各点は、取りうる可能性のある水素結合の組み合わせ(水素結合ネットワーク)。
 右下に位置するほど、よくある水素結合様式(network)を取っている。(Most likely networks)

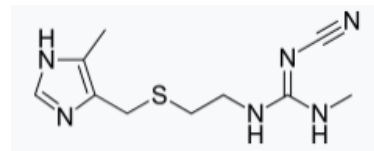


★Mean Co-ordination Score (y軸)とは、水素結合能力の平均。例えば、▼の場合、N1が2回HBしているのが0.031→2で割る
 Cl1が1回HBしているのは0.026
 O1が1回HBしているのが0.743
 これら3つの値を足し合わせ、平均をとった値をプロットしている。

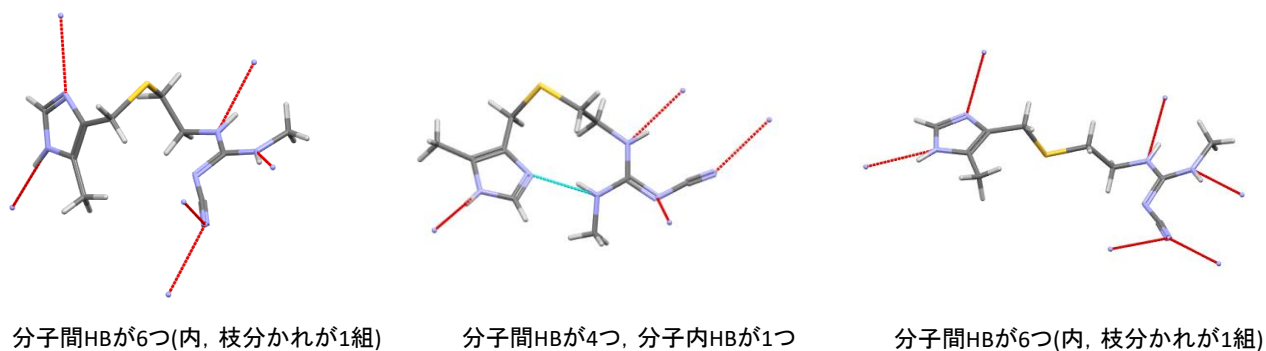
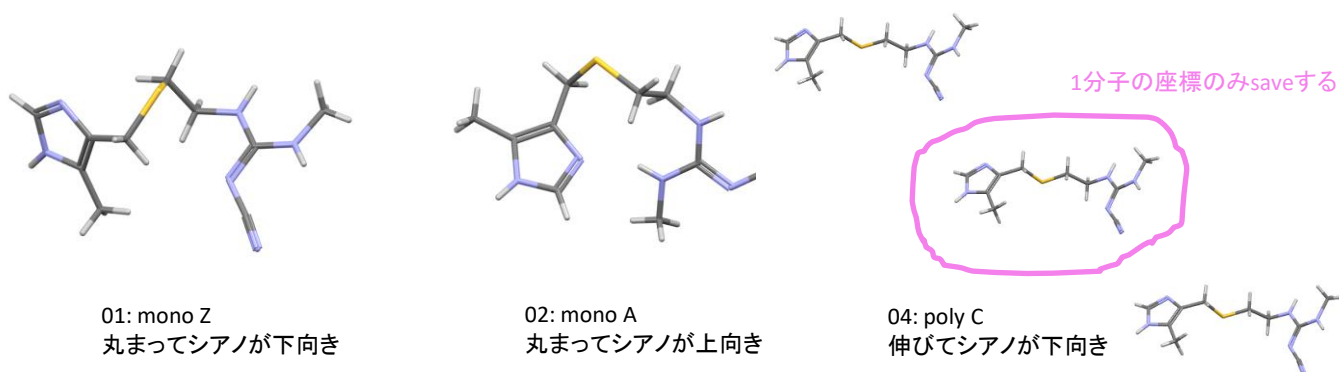
【例 2】 多型結晶が知られている系の場合

出典: Wikipedia

目的: Cimetidineは、ヒスタミンH2受容体拮抗剤 (H2ブロッカー)の一つとして知られている。H2受容体と拮抗することにより胃酸分泌を抑制することから胃酸抑制剤として使用される。このCimetidineには多型結晶があることが知られており、Hydrogen Bond Propensity toolにより、よりoptimalな構造について説明できるか、検証する。



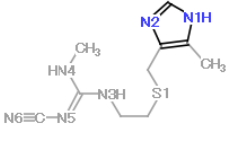
1. Cimetidineの多型結晶について、考察する。CSDのデータベースからCIMETD01(mono Z, P21/c), CIMETD02(mono A, P21/n), CIMETD04(polymorph C, C2/c)について確認する。Mercuryで表示すると、3つの結晶多型中の分子構造は異なった特徴を持つ。特にCIMETD04は、非対称単位が3分子で、それぞれはほぼ同じで伸びた分子構造を持つ。Edit→Auto Edit Structure...のManual Editの機能を使い、不要なmoleculeをRemoveし(座標が削除される), File→Save as..で.mol2などのフォーマットで構造ファイルを保存する。CIMETD01, 02についても、mol2で保存する。



水素結合ネットワークとしては似ている。
実験事実として、CIMETD04(針状)の結晶は得られにくい
ことがわかっている。

2. 【例1】同様、CIMETD01 に対し、1～6を行い、HBPチャートを計算する。

Donors and acceptors



Donors: N1, N3, N4
Acceptors: S1, N2, N5, N6

Functional groups

- Matched from library:
 - acyclic_n
 - al_al_thioether_1
 - al_amine_1
 - imidazole_2
 - nitrile

✓ All donors and acceptors matched

Generate Fitting Data

Auto generate fitting data structures

Generate Stop 100%

765 structures in fitting data (good size)

Analyse Cancel

Truncate data generation at #items 2000

Start analysis automatically

Use the slider to obtain sufficient and even group representation

Group	Count	Advice
1 acyclic_n	244	maybe sufficient
2 al_al_thioether_1	231	maybe sufficient
3 al_amine_1	494	good number
4 imidazole_2	231	maybe sufficient
5 nitrile	249	maybe sufficient

765 structures in fitting data (good size)

Analyse Cancel 100%

Analysis complete. Press 'Fit Model'.

Category	Label	# True	# False	Ignore?
1 Donor(s)	atom_0_of_al_amine_1 (matches N3,N4)	660	508	<input type="checkbox"/>
2	atom_3_of_imidazole_2 (matches N1)	270	227	<input type="checkbox"/>
3 Acceptor(s)	atom_0_of_acyclic_n (matches N5)	126	207	<input type="checkbox"/>
4	atom_0_of_al_al_thioether_1 (matches S1)	15	273	<input type="checkbox"/>
5	atom_0_of_imidazole_2 (matches N2)	227	214	<input type="checkbox"/>
6	atom_0_of_nitrile (matches N6)	165	193	<input type="checkbox"/>

Model Fitting

Use this page to **fit**, **assess** and **refine** a hydrogen bond logit model.

Refine Model...

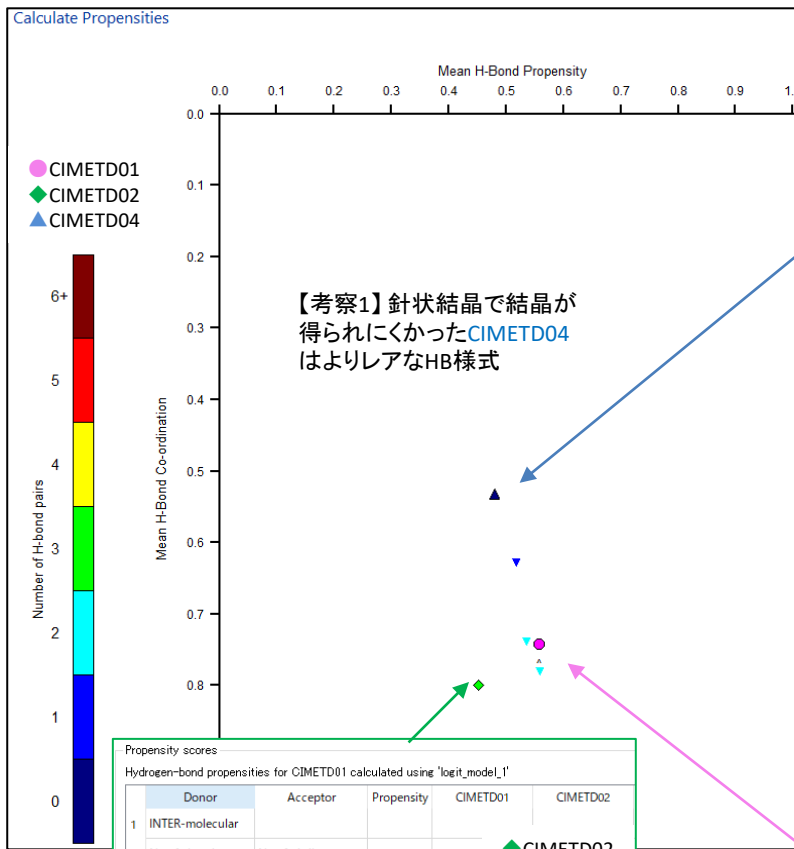
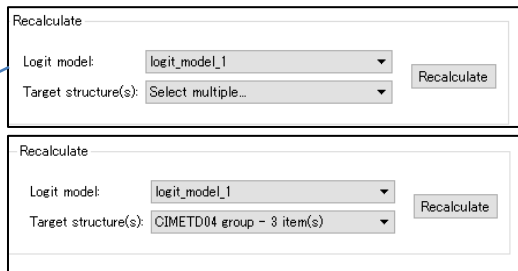
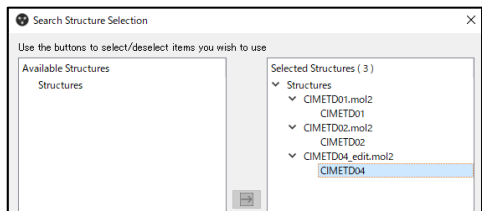
Model Coefficient Statistics

logit_model_1 Coefficients:

Coefficients:	Estimate	Std. Error	z value	Pr(> z)	Significance code	Lower Bound	Upper Bound
(Intercept)	1.696	0.290	5.852	4.84304e-9	***	1.130	2.267
Donoratom_3_of_imidazole_2	-0.330	0.150	-2.200	0.027782	*	-0.623	-0.036
Donorother	0.254	0.114	2.231	0.0256843	*	0.031	0.477
Acceptoratom_0_of_al_al_thioether_1	-2.134	0.333	-6.402	1.53294e-10	***	-2.832	-1.515
Acceptoratom_0_of_imidazole_2	0.848	0.191	4.435	9.18984e-6	***	0.476	1.227
Acceptoratom_0_of_nitrile	0.335	0.219	1.528	0.126544		-0.093	0.767
Acceptorother	1.740	0.166	10.495	9.12115e-26	***	1.419	2.069
Competition	-0.032	0.011	-2.936	0.00332965	**	-0.053	-0.011
Donor_steric_density	-0.019	0.003	-6.617	3.67611e-11	***	-0.025	-0.013
Acceptor_steric_density	-0.020	0.003	-5.820	5.86864e-9	***	-0.026	-0.013
Donor_aromaticity	-1.497	0.344	-4.351	1.35569e-5	***	-2.174	-0.824
Acceptor_aromaticity	-0.183	0.334	-0.548	0.583622		-0.838	0.474
Donoratom_0_of_al_amine_1	0.000	N/A	N/A	N/A	N/A	N/A	N/A
Acceptoratom_0_of_acyclic_n	0.000	N/A	N/A	N/A	N/A	N/A	N/A

Area under ROC curve = 0.811521 (good discrimination)

7. 多型結晶を同じチャートに表示したいので, RecalculateのSelect multipleから, 該当する3つのmol2ファイルを読み込み, Recalculateをクリックする.



Propensity scores
Hydrogen-bond propensities for CIMETD01 calculated using 'logit_model_1'

	Donor	Acceptor	Propensity	CIMETD01	CIMETD02	CIMETD04
1	INTER-molecular					
2	N4 of al_amine_1	N6 of nitrile	0.60			▲ CIMETD04
3	N4 of al_amine_1	N2 of imidazole_2	0.60	observed		
4	N3 of al_amine_1	N6 of nitrile	0.55	observed	observed	
5	N3 of al_amine_1	N2 of imidazole_2	0.55	observed		observed
6	N1 of imidazole_2	N6 of nitrile	0.52	observed		observed
7	N1 of imidazole_2	N2 of imidazole_2	0.52			
8	N4 of al_amine_1	N5 of acyclic_n	0.42			
9	N3 of al_amine_1	N5 of acyclic_n	0.37			
10	N1 of imidazole_2	N5 of acyclic_n	0.34		observed	observed
11	N4 of al_amine_1	S1 of al_al_thioeth...	0.06			

Left-clicking table items visualizes H-bond groups/contacts and interacts with chart

Atom Highlighting...

Co-ordination scores
(To refresh table: left-click chart point)

Atom (D/A)	= 0	= 1	= 2
1 N1 of sec_amin...	0.056	0.849	0.095
2 N3 of sec_amin...	0.269	0.716	0.015
3 N4 of sec_amin...	0.171	0.801	0.028
4 N2 of cyclic_n (a)	0.044	0.898	0.058
5 N5 of acyclic_n ...	0.804	0.196	0.000
6 N6 of nitrile (a)	0.274	0.683	0.043
7 S1 of acyclic_th...	0.966	0.034	0.000

Propensity scores
Hydrogen-bond propensities for CIMETD01 calculated using 'logit_model_1'

	Donor	Acceptor	Propensity	CIMETD01	CIMETD02
1	INTER-molecular				
2	N4 of al_amine_1	N6 of nitrile	0.60		◆ CIMETD02
3	N4 of al_amine_1	N2 of imidazole_2	0.60	observed	
4	N3 of al_amine_1	N6 of nitrile	0.55	observed	observed
5	N3 of al_amine_1	N2 of imidazole_2	0.55		
6	N1 of imidazole_2	N6 of nitrile	0.52	observed	
7	N1 of imidazole_2	N2 of imidazole_2	0.52		
8	N4 of al_amine_1	N5 of acyclic_n	0.42		
9	N3 of al_amine_1	N5 of acyclic_n	0.37		
10	N1 of imidazole_2	N5 of acyclic_n	0.34		observed
11	N4 of al_amine_1	S1 of al_al_thioeth...	0.06		

Left-clicking table items visualizes H-bond groups/contacts and interacts with chart

Atom

Propensity scores
Hydrogen-bond propensities for CIMETD01 calculated using 'logit_model_1'

	Donor	Acceptor	Propensity	CIMETD01
1	INTER-molecular			
2	N4 of al_amine_1	N6 of nitrile	0.60	● CIMETD01
3	N4 of al_amine_1	N2 of imidazole_2	0.60	observed
4	N3 of al_amine_1	N6 of nitrile	0.55	observed
5	N3 of al_amine_1	N2 of imidazole_2	0.55	
6	N1 of imidazole_2	N6 of nitrile	0.52	observed
7	N1 of imidazole_2	N2 of imidazole_2	0.52	
8	N4 of al_amine_1	N5 of acyclic_n	0.42	
9	N3 of al_amine_1	N5 of acyclic_n	0.37	
10	N1 of imidazole_2	N5 of acyclic_n	0.34	
11	N4 of al_amine_1	S1 of al_al_thioeth...	0.06	

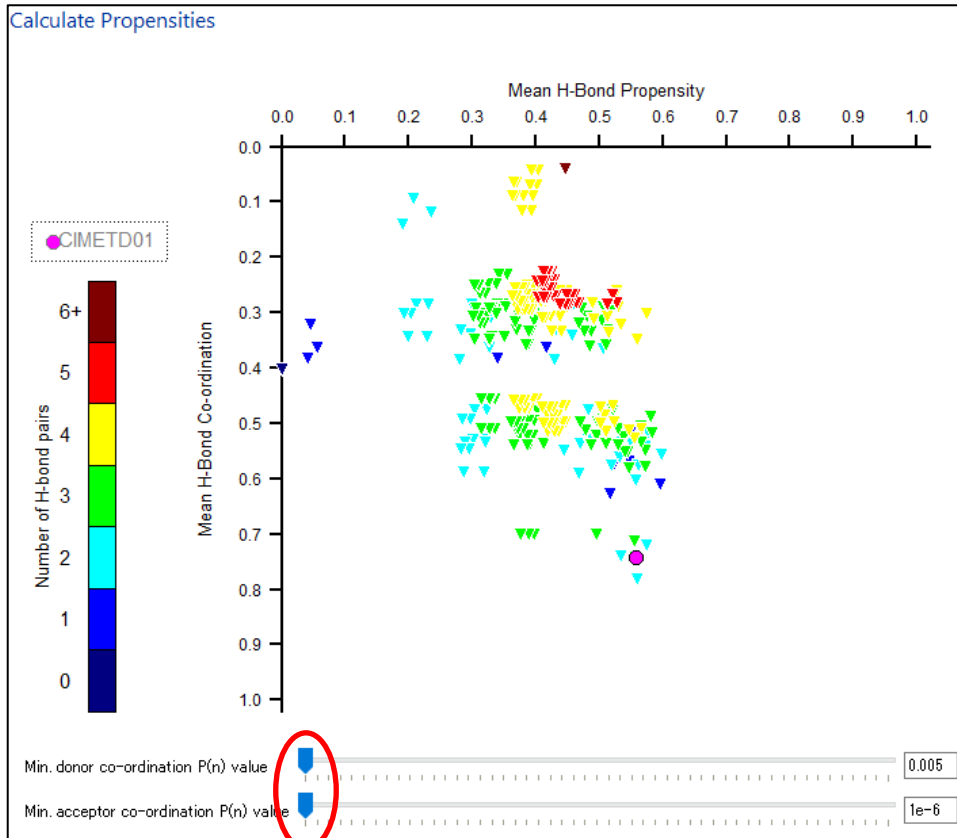
Left-clicking table items visualizes H-bond groups/contacts and interacts with chart

Co-ordination scores
(To refresh table: left-click chart point)

Atom (D/A)	= 0	= 1	= 2
1 N1 of sec_amin...	0.063	0.856	0.082
2 N3 of sec_amin...	0.216	0.765	0.020
3 N4 of sec_amin...	0.875	0.123	0.002
4 N2 of cyclic_n (a)	0.963	0.036	0.001
5 N5 of acyclic_n ...	0.585	0.415	0.000
6 N6 of nitrile (a)	0.192	0.758	0.050
7 S1 of acyclic_th...	0.966	0.034	0.000

【考察2】 CIMETD01は02よりもHBPスコアが大, Co-ordinationスコアは02の方が大きい, 実際のところ, 水素結合している原子(=1)のスコアが低い. 総合的には01の方が安定であることが推測される.

8. 【追加説明】Co-ordination scoreのminを下げるべく、スライダーを左に移動すると、隠れていた可能性のあるHBネットワークが表示される。色分けは、HBのペア数。
→分子間のHB数が極端に少なかったり、多かったりする系は結晶としてoptimalでないと思像できる。



このバーを左にスライドすると、より可能性の低いHBネットワークが表示される。

以上