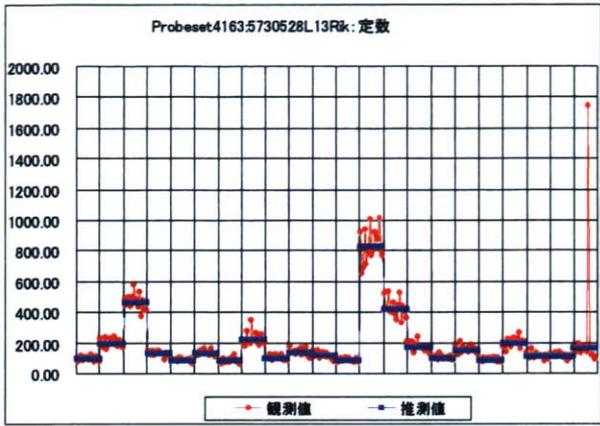


1.4.7. Langmuir式フィッティング検証 極低発現遺伝子

極低発現遺伝子:4163: 5730528L13Rik : 定数



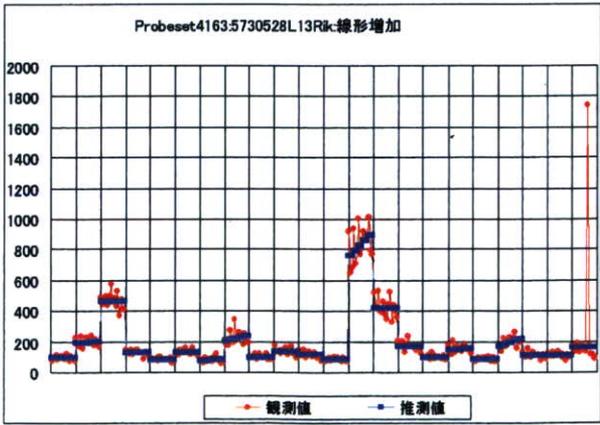
最もAICの小さい最適モデル

AIC=715.1

Copyright © NTT COMWARE 2008

1.4.7. Langmuir式フィッティング検証 極低発現遺伝子

極低発現遺伝子:4163: 5730528L13Rik : 線形(増加)

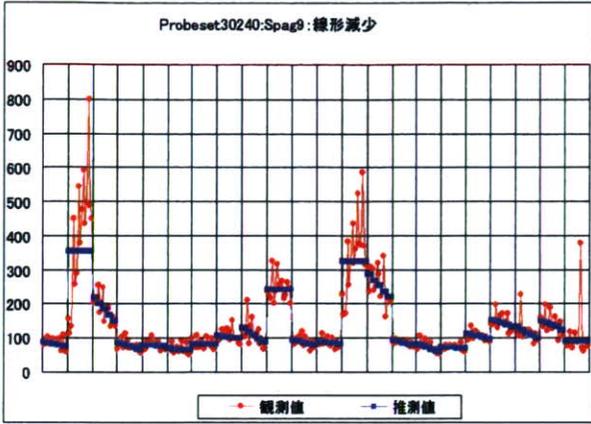


AIC=756.3

Copyright © NTT COMWARE 2008

1.4.7. Langmuir式フィッティング検証 極低発現遺伝子

極低発現遺伝子:30240:Spag9:線形減少

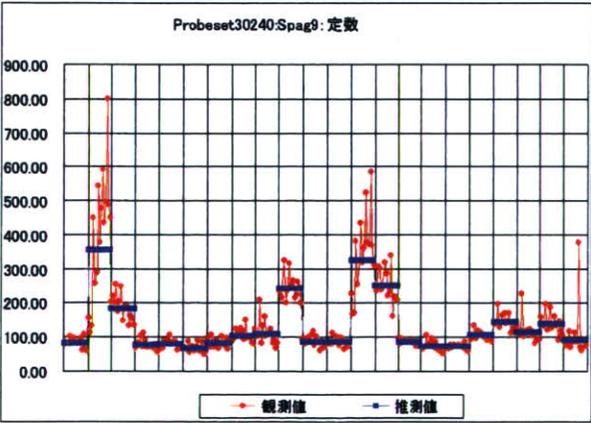


AIC=788.2

Copyright © NTT COMWARE 2008

1.4.7. Langmuir式フィッティング検証 極低発現遺伝子

極低発現遺伝子:30240:Spag9:定数

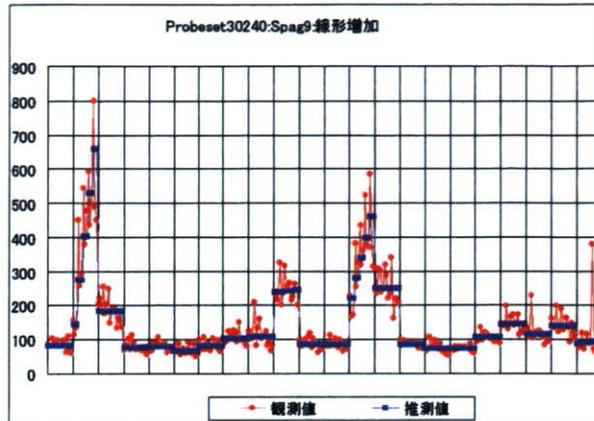


AIC=769.8

Copyright © NTT COMWARE 2008

1.4.7. Langmuir式フィッティング検証 極低発現遺伝子

極低発現遺伝子:30240:Spag9:線形(増加)



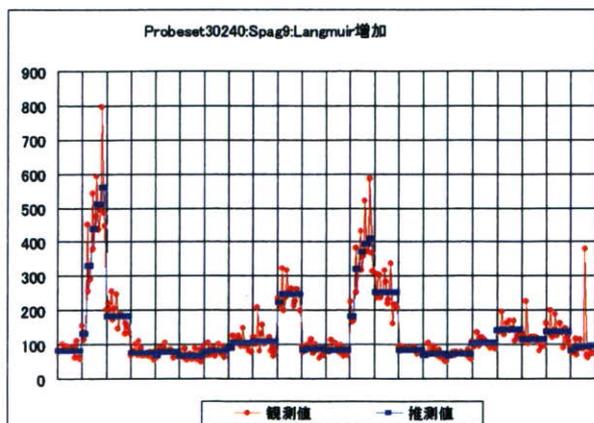
最もAICの小さい最適モデル

AIC=744.1

Copyright © NTT COMWARE 2008

1.4.7. Langmuir式フィッティング検証 極低発現遺伝子

極低発現遺伝子:30240:Spag9:Langmuir(増加)

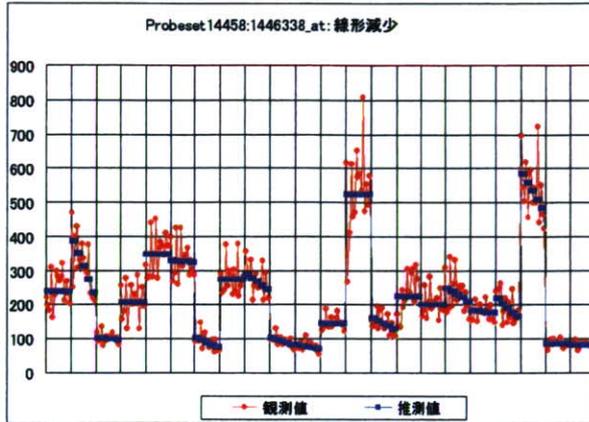


AIC=782.2

Copyright © NTT COMWARE 2008

1.4.7. Langmuir式フィッティング検証 極低発現遺伝子

極低発現遺伝子:14458:1446338_at:線形(減少)

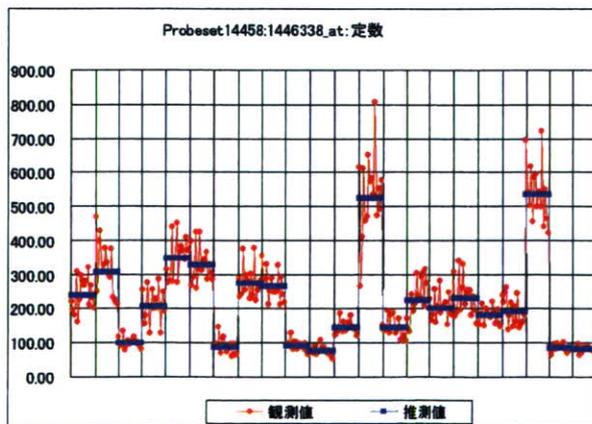


AIC=711.1

Copyright © NTT COMWARE 2008

1.4.7. Langmuir式フィッティング検証 極低発現遺伝子

極低発現遺伝子:14458:1446338_at:定数



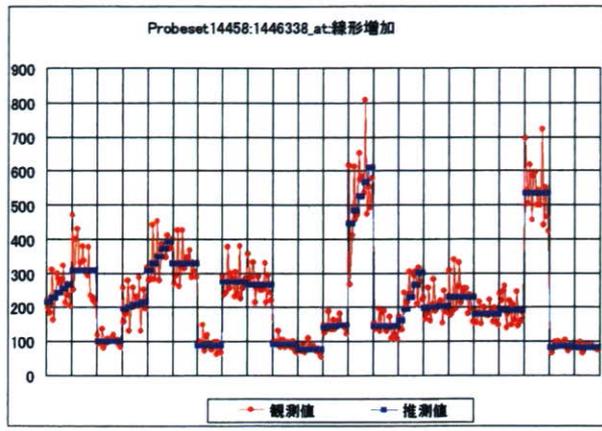
AIC=683.6

最もAICの小さい最適モデル

Copyright © NTT COMWARE 2008

1.4.7. Langmuir式フィッティング検証 極低発現遺伝子

極低発現遺伝子:14458:1446338_at:線形(増加)

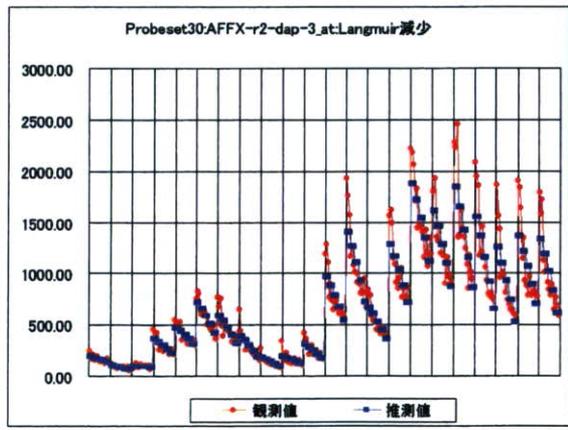


AIC=712.4

Copyright © NTT COMWARE 2008

1.4.8. Langmuir式フィッティング検証 Percellome Spika

Percellome Spike:30:AFFX-r2-Bs-dap-3_at:Langmuir(減少)

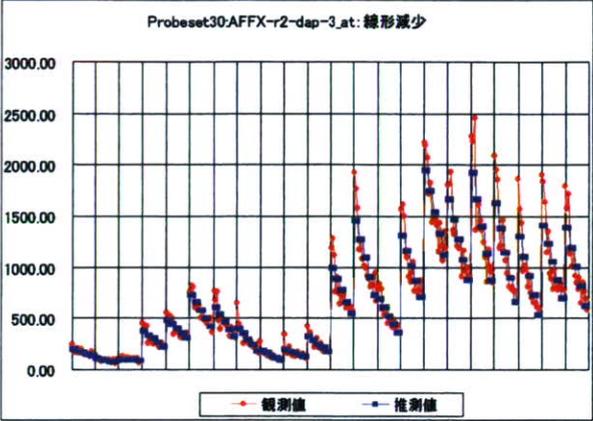


AIC=723.0

Copyright © NTT COMWARE 2008

1.4.8. Langmuir式フィッティング検証 Percellome Spike...

Percellome Spike: 30: AFFX-r2-Bs-dap-3_at: 線形(減少)



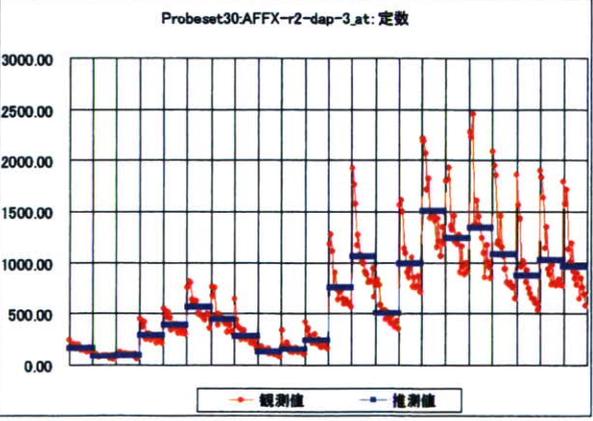
最もAICの小さい最適モデル

AIC=663.2

Copyright © NTT COMWARE 2008

1.4.8. Langmuir式フィッティング検証 Percellome Spike...

Percellome Spike: 30: AFFX-r2-Bs-dap-3_at: 定数

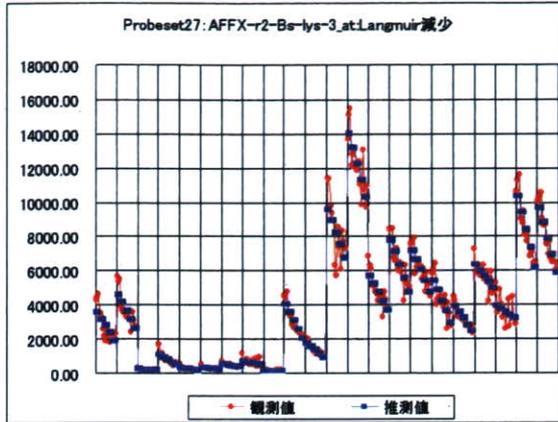


AIC=834.3

Copyright © NTT COMWARE 2008

1.4.8. Langmuir式フィッティング検証 Percellome Spike

Percellome Spike:27:AFFX-r2-Bs-lys-3_at:Langmuir (減少)

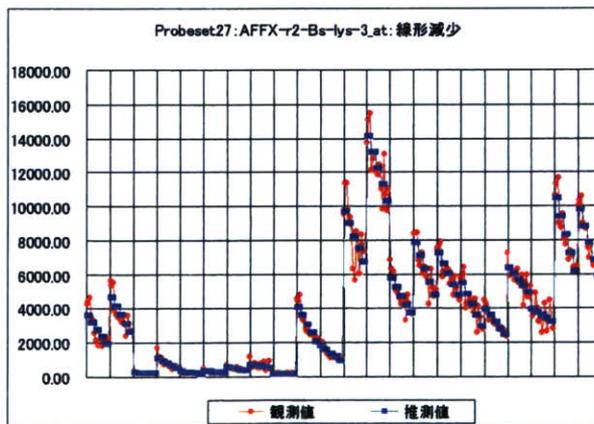


AIC=718.7

Copyright © NTT COMWARE 2008

1.4.8. Langmuir式フィッティング検証 Percellome Spike

Percellome Spike:27:AFFX-r2-Bs-lys-3_at:Langmuir (減少)



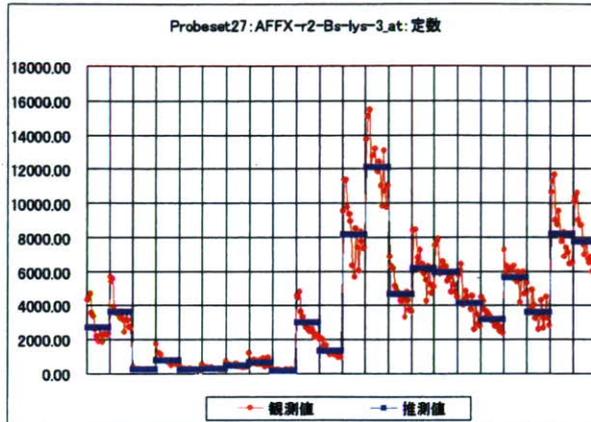
最もAICの小さい最適モデル

AIC=670.2

Copyright © NTT COMWARE 2008

1.4.8. Langmuir式フィッティング検証 Percellome Spika...

Percellome Spike:27:AFFX-r2-Bs-lys-3_at:定数

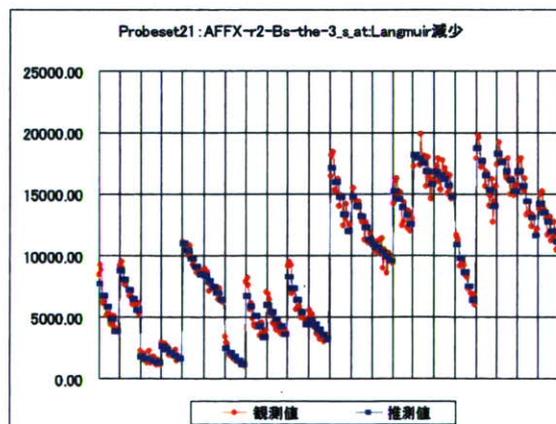


AIC=768.8

Copyright © NTT COMWARE 2008

1.4.8. Langmuir式フィッティング検証 Percellome Spika...

Percellome Spike:21:AFFX-r2-Bs-thr-3_s_at:Langmuir (減少)

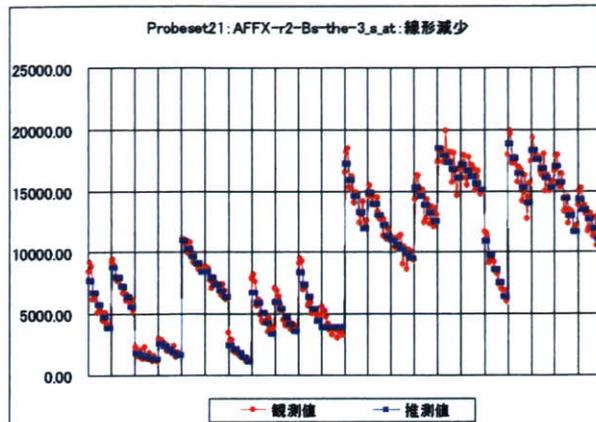


AIC=614.3

Copyright © NTT COMWARE 2008

1.4.8. Langmuir式フィッティング検証 Percellome Spika...

Percellome Spike: 21: AFX-r2-Bs-thr-3_s_at: 線形(減少)



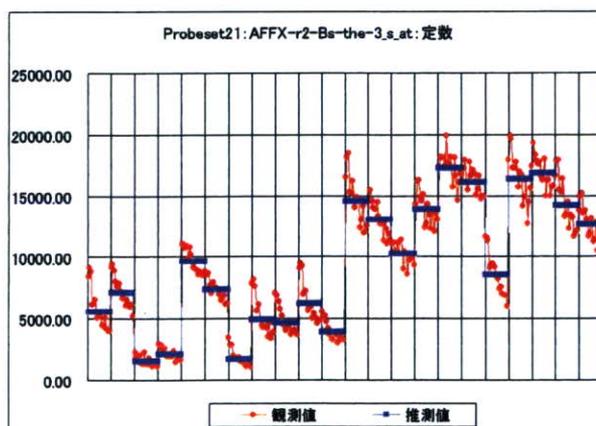
最もAICの小さい最適モデル

AIC=576.2

Copyright © NTT COMWARE 2008

1.4.8. Langmuir式フィッティング検証 Percellome Spika...

Percellome Spike: 21: AFX-r2-Bs-thr-3_s_at: 定数



AIC=675.0

Copyright © NTT COMWARE 2008

1.4.9. Langmuir式フィッティング検証 AICによる選択まとめ

- 高発現域では、Liver側飽和
- 中発現域、Spikeでは、Liver側が少ない線形モデル
 - Brainの細胞あたりmRNA発現量が少ないのが反映していると考えられる
- 低発現域では、定数モデルが選択された

最小AIC

	No	PinNo	Avg	SD	AffID	GeneSymbol	CV	Langmuir 係数推定	線形モ デル	定数モ デル	線形モ デル	Langmuir 係数推定
飽和	0.18	48	559	7700.28	68.10	1460258.st	0.81			2588.4	927.6	714.9
	0.28	88	12010	6695.40	230.26	1448792.st	0.39			2788.1	830.2	590.1
	0.18	60	43882	7389.56	278.33	1418809.st	0.49			2895.2	1076.8	672.9
線形	0.08	16	40381	8888.57	284.42	1420485.st	0.39			2885.0	1181.2	884.7
	0.08	2984	8363	777.05	27.05	1465454.st	0.39			1543.4	723.8	447.4
	38.11	17178	32281	9339	9.93	1455342.st	0.19			838.1	723.5	734.6
高発現域	0.08	1	42821	11035.10	223.28	1417850.st	0.28			2898.3	1083.0	659.3
	0.08	2	43215	10590.20	436.38	1417856.st	0.48			3103.6	988.5	619.0
	0.08	3	26886	10619.77	485.48	1438110.st	0.48			2671.6	1015.8	647.6
中発現域	0.08	4	5638	10504.17	388.18	1465201.st	0.39			2887.8	1020.1	678.5
	0.08	5	41538	10193.04	234.14	1419233.st	0.28			2850.7	1016.8	590.8
	10.08	4613	45083	410.25	48.31	1415708.st	0.19	738.5	672.8			
低発現域	10.08	4614	817	470.16	10.17	1458880.st	0.28	738.5	672.8			
	10.08	4615	38943	489.78	15.78	1421853.st	0.39	688.6	607.6			
	10.08	4616	32355	469.60	88.78	1428441.st	0.19	690.9	641.0			
低発現域	10.08	4617	8029	469.59	4.72	1465748.st	0.19	691.4	1042.5			603.9
	50.08	22651	26924	49.24	4.78	1435974.st	0.19	800.3	754.4			811.4
	50.08	22652	8808	49.64	11.65	1464001.st	0.28		673.5		649.3	688.1
低発現域	50.08	22653	43844	49.62	11.89	1417127.st	0.28		681.2		627.6	656.5
	50.08	22654	16795	49.61	10.41	1448041.st	0.19	782.0	717.8		727.5	789.4
	50.08	22655	19740	49.61	11.62	1441055.st	0.28		688.4		699.8	699.7
低発現域	80.08	38021	28431	13.81	3.68	1422008.st	0.19		693.1		638.1	650.0
	80.08	38023	20455	13.47	2.67	1446311.st	0.19		685.6		637.8	680.0
	80.08	38024	4163	13.47	0.71	1468644.st	0.19		755.4		716.1	756.3
低発現域	80.08	38025	20240	13.47	4.07	1436566.st	0.19		788.2		769.8	744.1
	80.08	38026	14458	13.47	11.80	1446338.st	0.19		711.1		683.6	712.4
	80.08	38027	28581	13.47	4.68	1465748.st	0.19					
Percollome Scale	18.98	7828	3	265.81	4.68	1465748.st	0.19					
	6.68	2385	30	102.82	27.91	1465748.st	0.19	723.0	693.2		634.3	
	2.48	1089	27	1832.82	71.83	1465748.st	0.19					
低発現域	0.58	208	27	4901.25	362.87	1465748.st	0.19	718.7	670.2		788.8	
	0.08	16	21	8829.21	388.22	1465748.st	0.19	614.3	676.2		676.0	

Copyright © NTT COMWARE 2008

1.5.1. 全プローブ計算①最適化順序



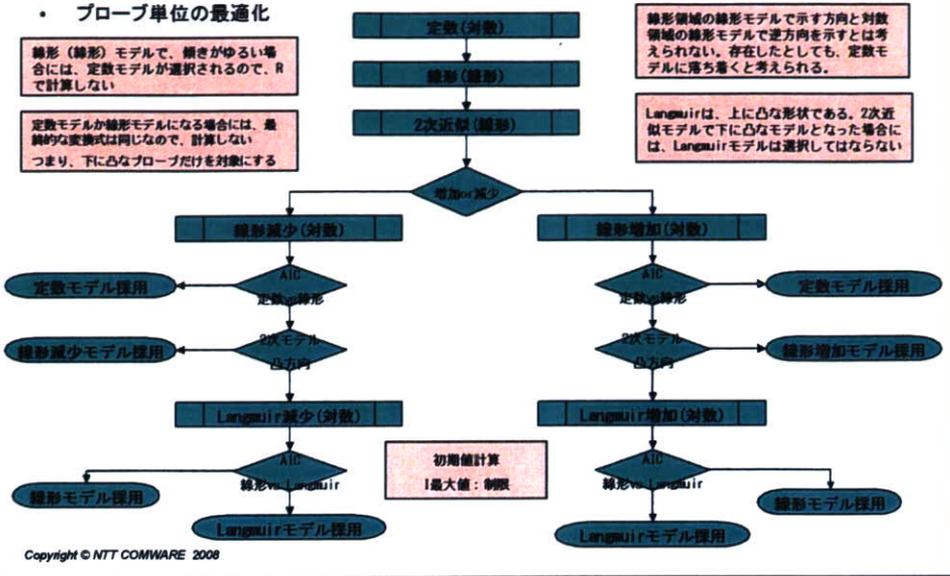
- Langmuirの係数を求めるための数値最適化アルゴリズムは、適切なモデルと初期値を与えてうまく解に誘導させる必要がある
 - 次の弊害が考えられる
 - 収束までに時間がかかる
 - 別の局所解で収束判定に引っかかる
- モデル選択のための簡易モデルで事前情報で振り分け
 - 線形領域の線形モデル vs. 対数領域の線形モデル
 - 線形領域の線形モデルで示す方向と対数領域の線形モデルで逆方向を示すとは考えられない。存在したとしても、定数モデルに落ち着くと考えられる
 - 2次近似モデル
 - Langmuirは、上に凸な形状である。2次近似モデルで下に凸なモデルとなった場合には、Langmuirモデルは選択しない
- 収束判定チェック(追加項目)
 - モデルのパラメータ数が増えた場合には尤度は必ず小さくなる。小さくならないならば、適切な収束していないと判断できる

Copyright © NTT COMWARE 2008

1.5.1. 全プローブ計算①最適化順序



・ プローブ単位の最適化

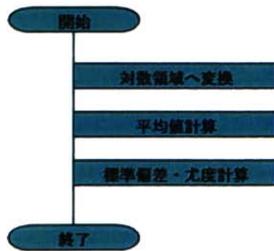


Copyright © NTT COMWARE 2008

1.5.2. 全プローブ計算①定数モデル(対数領域)



・ Teradataを使用し、SQL1文で計算を実施する



尤度

$$L_{n,n}(\mu, \sigma) = \prod_{i=1}^n L_n(\mu, \sigma) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_i - \mu)^2}{2\sigma^2}}$$

対数尤度

$$\begin{aligned} \log L_{n,n}(\mu, \sigma) &= \log \left(\prod_{i=1}^n L_n(\mu, \sigma) \right) = \log \left(\prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_i - \mu)^2}{2\sigma^2}} \right) \\ &= \sum_{i=1}^n \log \left(\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_i - \mu)^2}{2\sigma^2}} \right) \\ &= \sum_{i=1}^n \log \left(\frac{1}{\sqrt{2\pi}\sigma} \right) + \sum_{i=1}^n \log \left(e^{-\frac{(x_i - \mu)^2}{2\sigma^2}} \right) \\ &= -n \log \sqrt{2\pi}\sigma - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2 \end{aligned}$$

パラメータは、平均と標準偏差の2個

$$AIC = -2\log(L) + 2k$$

$$= 2n \log \sqrt{2\pi} + 2n \log \sigma + \frac{n}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2 + 2 \cdot 2$$

Copyright © NTT COMWARE 2008

1.5.3. 全プローブ計算①定数モデル(線形領域)



- Teradataを使用し、SQL1文で計算を実施する
 - Reqr_slope関数を用いる
- 対数尤度・AICは計算しない、傾きの正負を求める

```
select
  probeset_no, probno, pm,
  average(val), regr_slope(val, lbn_no)
from tbl_ce13 t1
join tbl_exp t2
  on t1.exp_no=t2.exp_no
group by probeset_no, pm, probno
order by probeset_no, pm, probno
```

Copyright © NTT COMWARE 2008

1.5.4. 全プローブ計算① 2次関数モデル(線形領域)



- Teradataを使用し、SQL1文で計算を実施する
 - 事前に3×3の逆行列を求めておくことにより、行列の積で2次関数モデルを求める
- 対数尤度・AICは計算しない、2次の係数の正負を求める

$$\begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \\ \vdots & \vdots & \vdots \\ 25 & 5 & 1 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} v_1 \\ v_2 \\ v_3 \\ \vdots \\ v_{15} \end{pmatrix}$$

最小2乗法

$$'(Kx-v)(Kx-v)$$

$$= '(Kx)(Kx) - 2'(v)(Kx) + 'vv$$

$$= 'x'KKx - 2'vKx + 'vv$$

微分すると

$$'KKx = 'vK$$

$$x = ('KK)^{-1} 'vK$$

$$\begin{pmatrix} 0.02381 & -0.14286 & 0.166667 \\ -0.14286 & 0.890476 & -1.1 \\ 0.166667 & -1.1 & 1.533333 \end{pmatrix}$$

```
Select
  probeset_no, pm, probno,
  sum(val) As v1,
  sum(val*lbn_no) As v2,
  sum(val*lbn_no*lbn_no) As v3,
  0.023809524*v3-0.142857143*v2          +0.16666667*v1 As coeff_a,
  -0.142857143*v3+0.89047619*v2        -1.1*v1 As coeff_b,
  0.16666667*v3-1.1*v2                  +1.53333333*v1 As coeff_c
Group   By 1,2,3
From    tbl_ce13 t1
Join    tbl_exp t2
       On t1.exp_no=t2.exp_no
Where   t1.probeset_no=1
Order   By 1,2,3
```

Copyright © NTT COMWARE 2008

1.5.5. 全プローブ計算① Teradataによる線形モデル計算

- Teradataを用いた線形計算をPM補正結果で実施した。
- 線形領域線形モデルの傾きが平均値の2%以上という条件で、減少型133,359プローブ、増加型131,367プローブ、計264,726プローブを実施する。

	Langmuir対象
減少	133,359
増加	131,367
総計	264,726

Copyright © NTT COMWARE 2008

1.5.5. 全プローブ計算① Rによる計算方法



- 各プローブに対して、非線形最適化の計算を行う必要がある。
- 次の工夫により、計算速度向上を図る
 - 線形モデル最適化後にLangmuir最適化を行う
 - 100万回の計算回数を25万程度に減らす
 - 1回Rを起動させて、1000probesetsの非線形最適化を実施するように工夫する。
 - 計算途中のRの起動回数を減らし、速度向上に努める

Copyright © NTT COMWARE 2008

1.5.6. 全プローブ計算① Langmuirモデル(対数領域)



- Rによる非線形最適化を行う
- 収束結果判定
 - 結果コード=2の場合、kpが大きすぎると失敗
 - Steptol=1.0e-17として、完全不動の状態でしか収束させない
- 初期値
 - Ipは、最大値と最小値の差を与える
 - 増加の場合、100:0の値に100:0と25:75の差を加えて、最大値のおおよその値を見つける
 - Bgpは、最小値の半分

Copyright © NTT COMWARE 2008

1.5.7. 全プローブ計算① Rスクリプト基本構造



- 線形モデル
 - 入力
 - 識別子:プローブセット番号、PM、プローブ番号
 - データ 15個
 - LBM比率はカラム名に
 - 線形領域計算結果
 - 傾き、切片
 - 初期値
 - 線形領域計算を用いる
 - AIC計算:
 - 定数モデルとAIC判別
 - 標準偏差は小さくなっているはず、なっていないければ、警告
- Langmuirモデル
 - 入力
 - 識別子:プローブセット番号、PM、プローブ番号
 - データ
 - 初期値:
 - データから作成
 - AIC計算
 - 線形モデルとAIC判別
 - 標準偏差は小さくなっているはず、なっていないければ警告

Copyright © NTT COMWARE 2008

1.5.8. 全プローブ計算① Rスクリプト生成&R結果ファイル整形プログラム



- Teradataの出力結果を元にRスクリプトを作成するプログラム
 - Visual C++で作成する
- Rスクリプトの実行結果をTeradataへロード可能なファイルにするための整形プログラム
 - AWKで作成する

Copyright © NTT COMWARE 2008

1.5.9. 全プローブ計算①計算結果 (Cyp7a1)



- Cyp7a1に着目してプローブ単位で確認
 - QPCRにてサーフェスの作成がされているので、確認を行いやすいので、このプローブセットを用いた。
- Langmuirモデルが選択されたのは、22プローブ中12プローブ (MM3個、PM9個)

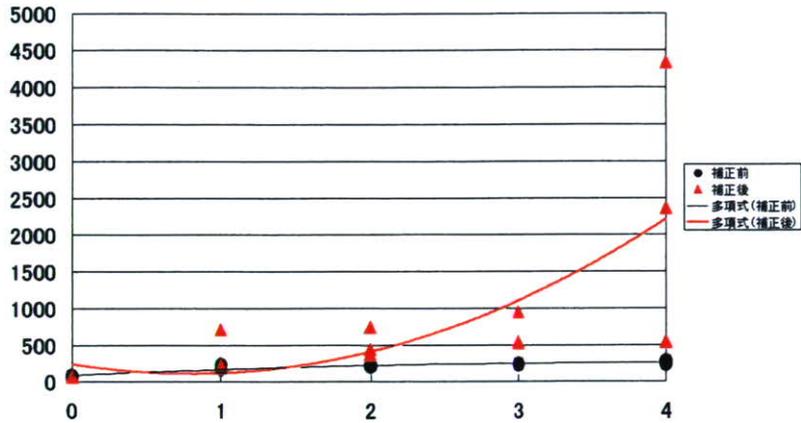
	ip	bx	Linear				Langmuir				diff		
			code	iteration	SD	AIC	code	iteration	SD	AIC	SD	AIC	
0													
1													
2													
3													
4	225.1261	75.1810	1	9	0.2075	93.39	2	32	0.1291	81.18	0.0783	12.23	
5	174.0594	107.9360	1	8	0.1804	85.67	1	28	0.1482	85.29	0.0123	0.39	
6													
7													
8	172.5238	38.6229	1	10	0.1394	81.47	1	50	0.1183	78.53	0.0211	2.93	
9													
10													
0													
1	380.3365	84.8778	1	8	0.2131	94.19	1	47	0.1891	92.61	0.0240	1.58	
2													
3	630.3905	83.3525	1	8	0.1858	90.08	1	42	0.1496	85.58	0.0362	4.50	
4	1242.9140	50.8938	1	11	0.1957	91.84	1	66	0.1208	79.12	0.0751	12.52	
5	714.4125	105.1412	1	8	0.1378	81.11	1	31	0.1113	76.71	0.0285	4.40	
6	631.9981	29.9714	1	13	0.1992	92.18	1	55	0.1295	81.25	0.0697	10.91	
7	1309.7798	154.8640	1	10	0.1694	87.30	1	34	0.1180	78.47	0.0513	8.83	
8	1138.6972	110.5632	2	16	0.1676	86.99	2	32	0.1033	74.47	0.0643	12.51	
9	301.8765	166.5667	1	9	0.1461	82.86	1	34	0.1324	81.91	0.0137	0.95	
10	867.9886	188.3824	1	15	0.1564	84.91	1	32	0.1241	79.98	0.0322	4.93	

Copyright © NTT COMWARE 2008

1.5.9. 全プローブ計算①計算結果 (Cyp7a1)



Cyp7a1 MM Probe4



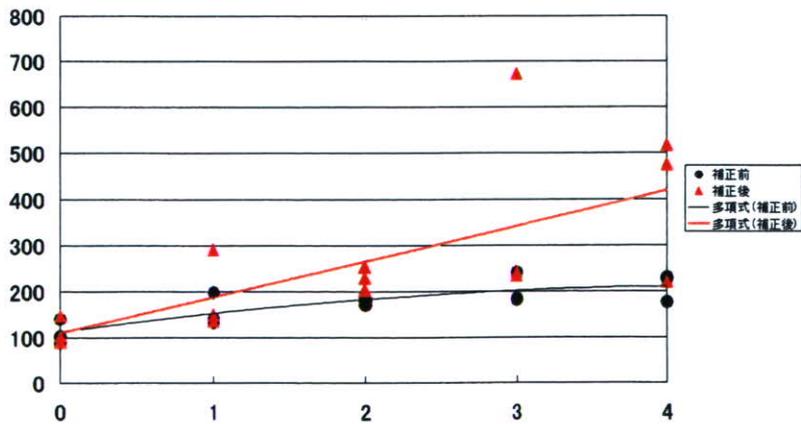
補正後でLiver側で大きな値に変換されている

Copyright © NTT COMWARE 2008

1.5.9. 全プローブ計算①計算結果 (Cyp7a1)



Cyp7a1 MM Probe5



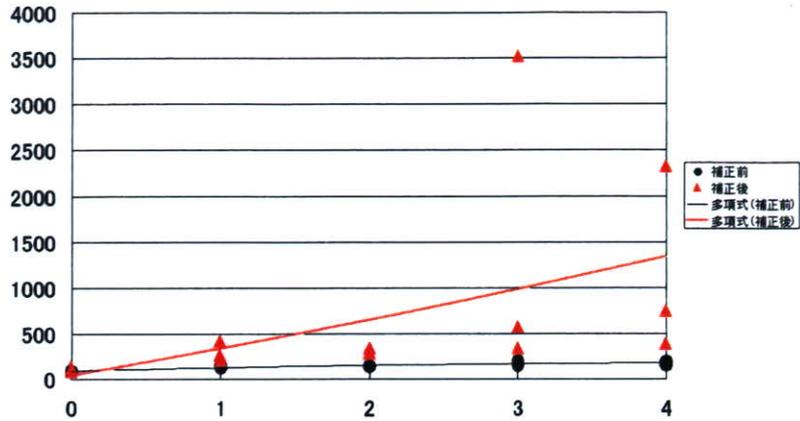
補正がうまくいっているように見える

Copyright © NTT COMWARE 2008

1.5.9. 全プローブ計算①計算結果 (Cyp7a1)



Cyp7a1 MM Probe8



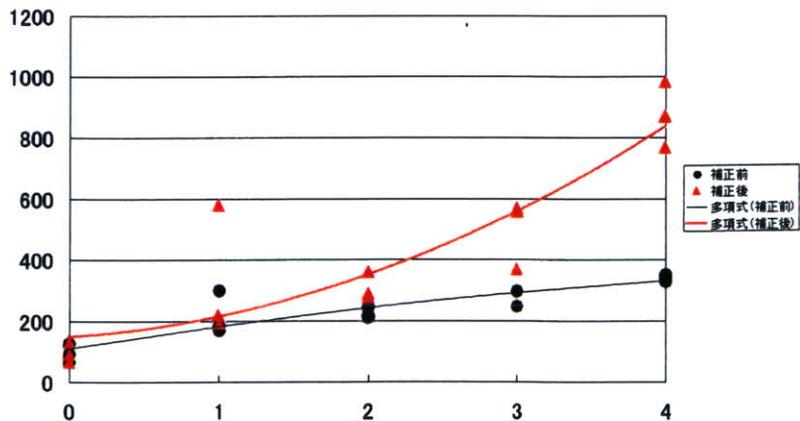
補正がうまくいっているように見える

Copyright © NTT COMWARE 2008

1.5.9. 全プローブ計算①計算結果 (Cyp7a1)



Cyp7a1 PM Probe1



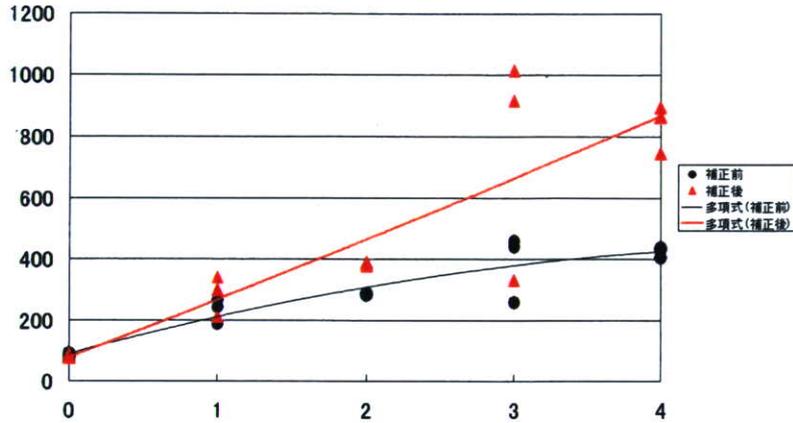
補正がうまくいっているように見える

Copyright © NTT COMWARE 2008

1.5.9. 全プローブ計算①計算結果 (Cyp7a1)



Cyp7a1 PM Probe3



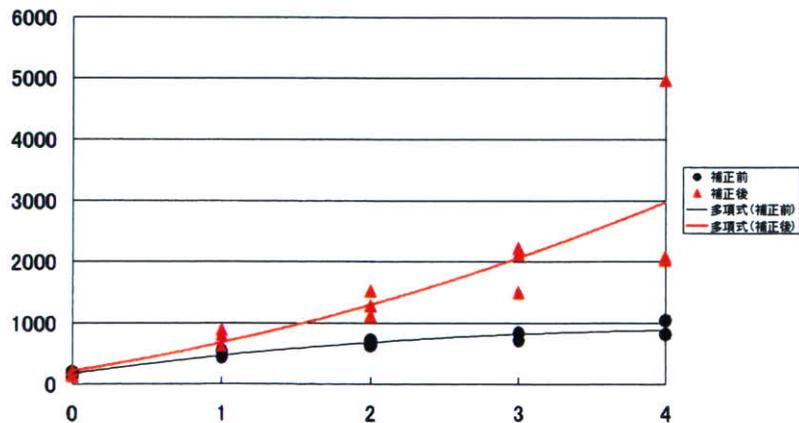
補正がうまくいっているように見える

Copyright © NTT COMWARE 2008

1.5.9. 全プローブ計算①計算結果 (Cyp7a1)



Cyp7a1 PM Probe4



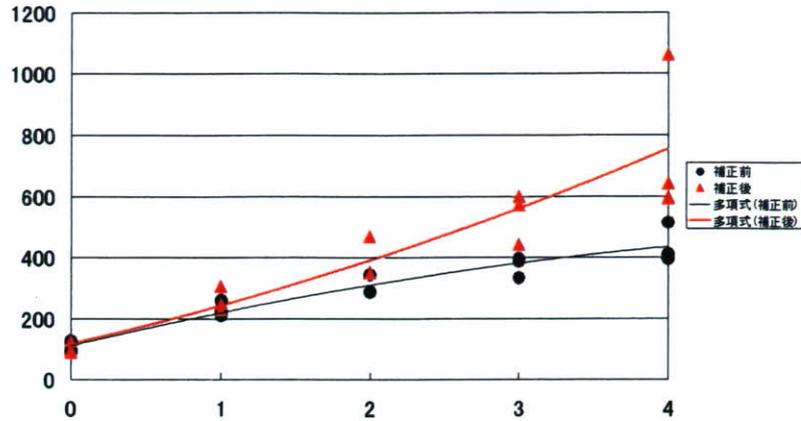
Probe4MM と同様に大きな値に変換されている

Copyright © NTT COMWARE 2008

1.5.9. 全プローブ計算①計算結果 (Cyp7a1)



Cyp7a1 PM Probe5



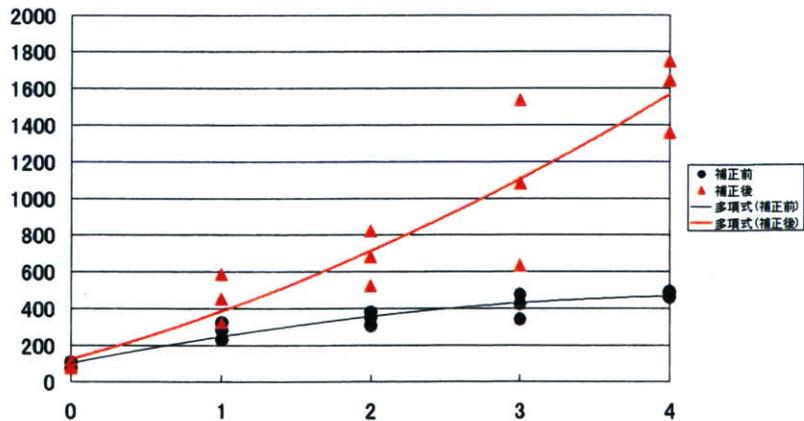
補正がうまくいっているように見える

Copyright © NTT COMWARE 2008

1.5.9. 全プローブ計算①計算結果 (Cyp7a1)



Cyp7a1 PM Probe6



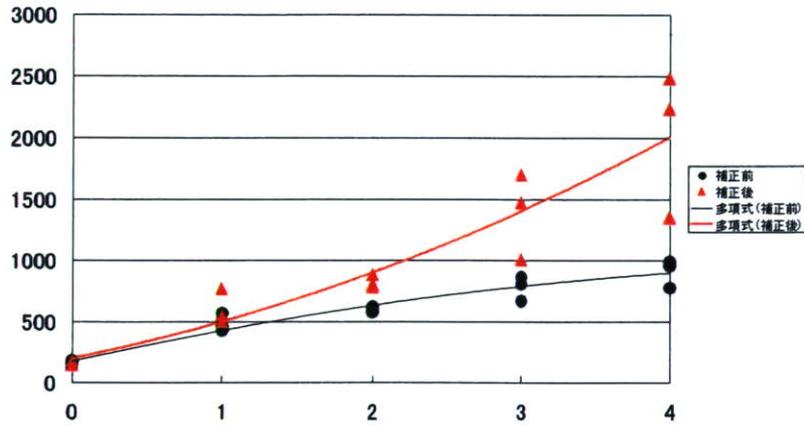
補正がうまくいっているように見える

Copyright © NTT COMWARE 2008

1.5.9. 全プローブ計算①計算結果 (Cyp7a1)



Cyp7a1 PM Probe7



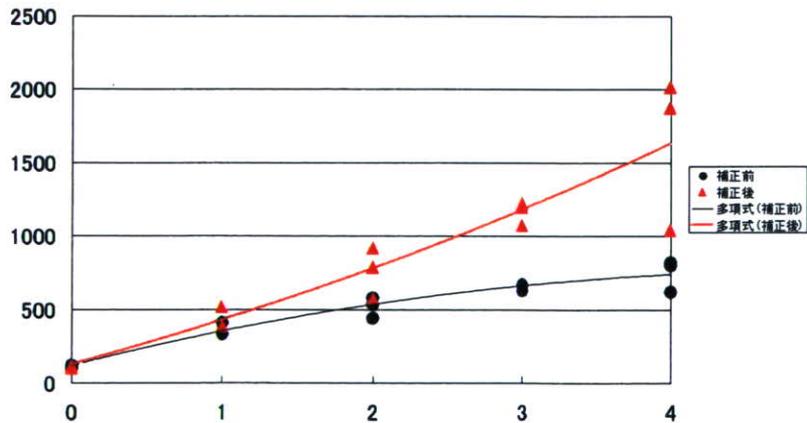
補正がうまくいっているように見える

Copyright © NTT COMWARE 2008

1.5.9. 全プローブ計算①計算結果 (Cyp7a1)



Cyp7a1 PM Probe8



補正がうまくいっているように見える

Copyright © NTT COMWARE 2008