Supplementary Information

Table S1: Fatty acid composition (% total lipids) of the oil extracted from *Thraustochytrium* sp. AH-2 (PRA-296^{TM}) by GC-FAME technique according to the protocol and conditions described in Ref. 8.

Fatty acids	**Common Name**	**Amount**[*a*] **(% total lipids)**
C14:0	myristic acid	14.91 ± 0.63
C15:0	pentadecylic acid	1.23 ± 0.13
C16:0	palmitic acid	24.90 ± 1.01
C17:0	margaric acid	0.21 ± 0.05
C18:0	stearic acid	0.94 ± 0.13
C22:0	behenic acid	0.42 ± 0.05
Sum SFAs		*42.61 ± 0.30*
C16:1	palmitoleic acid	2.88 ± 0.19
C18:1n-9	oleic acid	0.52 ± 0.24
C18:1	methyl *trans*-vaccenate	2.15 ± 0.73
C22:1n-9	erucic acid	0.54 ± 0.11
C24:1n-9	nervonic acid	0.07 ± 0.03
Sum MUFAs		*6.15 ± 0.44*
C18:3n-3	α-linolenic acid (ALA)	0.00 ± 0.00
C20:4n-3	eicosatetraenoic acid (ETA)	1.26 ± 0.05
C20:5n-3	eicosapentaenoic acid (EPA)	3.75 ± 0.09
C22:5n-3	docosapentaenoic acid (DPA)	1.63 ± 0.07
C22:6n-3	docosahexaenoic acid (DHA)	34.67 ± 2.07
C18:2n-6	linoleic acid (LA)	0.24 ± 0.01
C20:3n-6	dihomo-γ-linolenic acid (DGLA)	0.00 ± 0.00
C20:4n-6	arachidonic acid (AA)	0.00 ± 0.00
C22:5n-6	docosapentaenoic acid (osbond acid)	9.73 ± 0.42
Sum PUFAs		*51.29 ± 0.45*

[*a*] values are mean ± standard deviation obtained in triplicates.

Figure S2: Average EMSC-corrected (*a*) absorbance and (*b*) 2nd derivative spectra of dehydrated formalin-fixed (*blue*) and hydrated live (*red*) thraustochytrium cells (PRA-296^{TM}) acquired using FPA-FTIR and synchrotron FTIR microspectroscopic techniques, respectively.

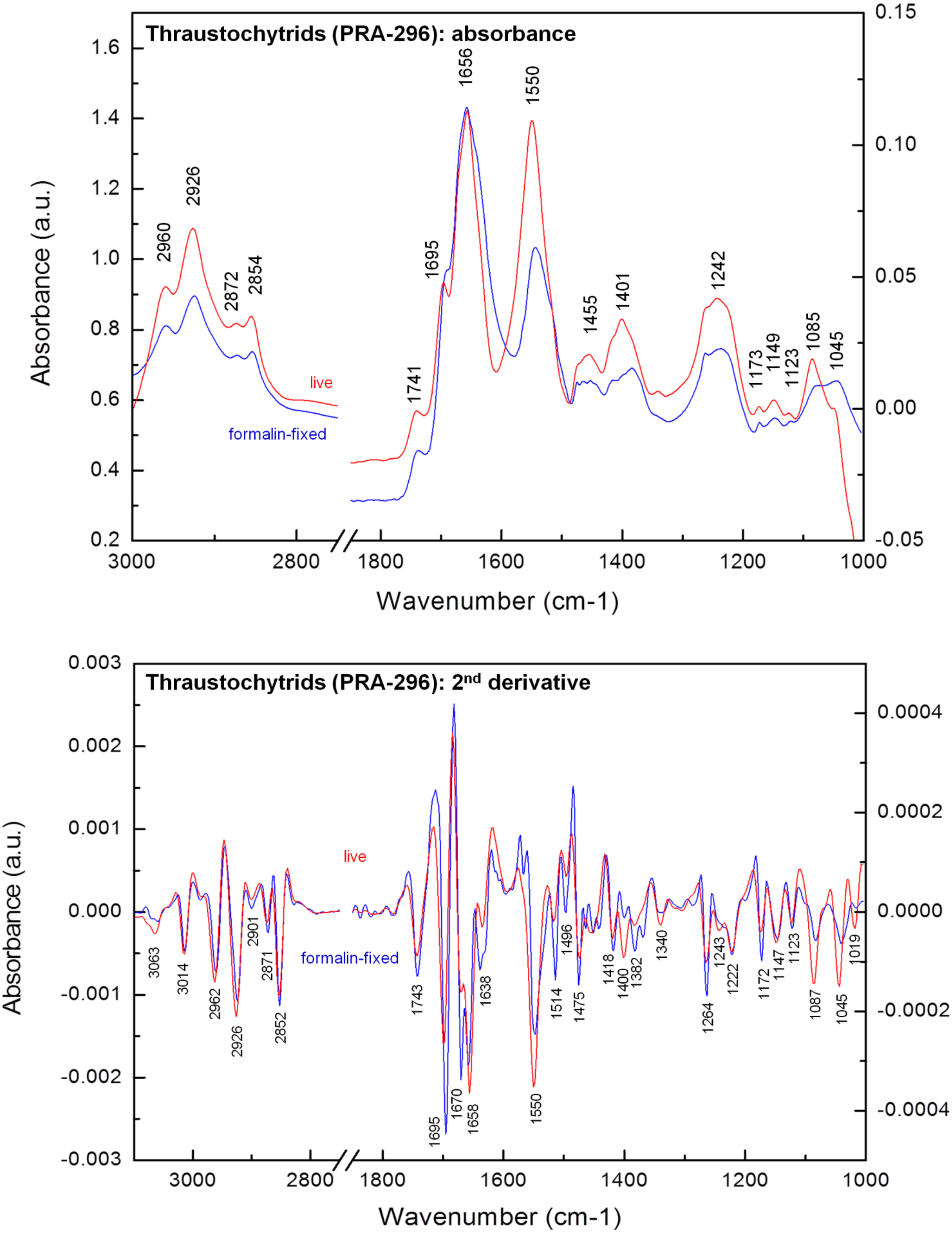
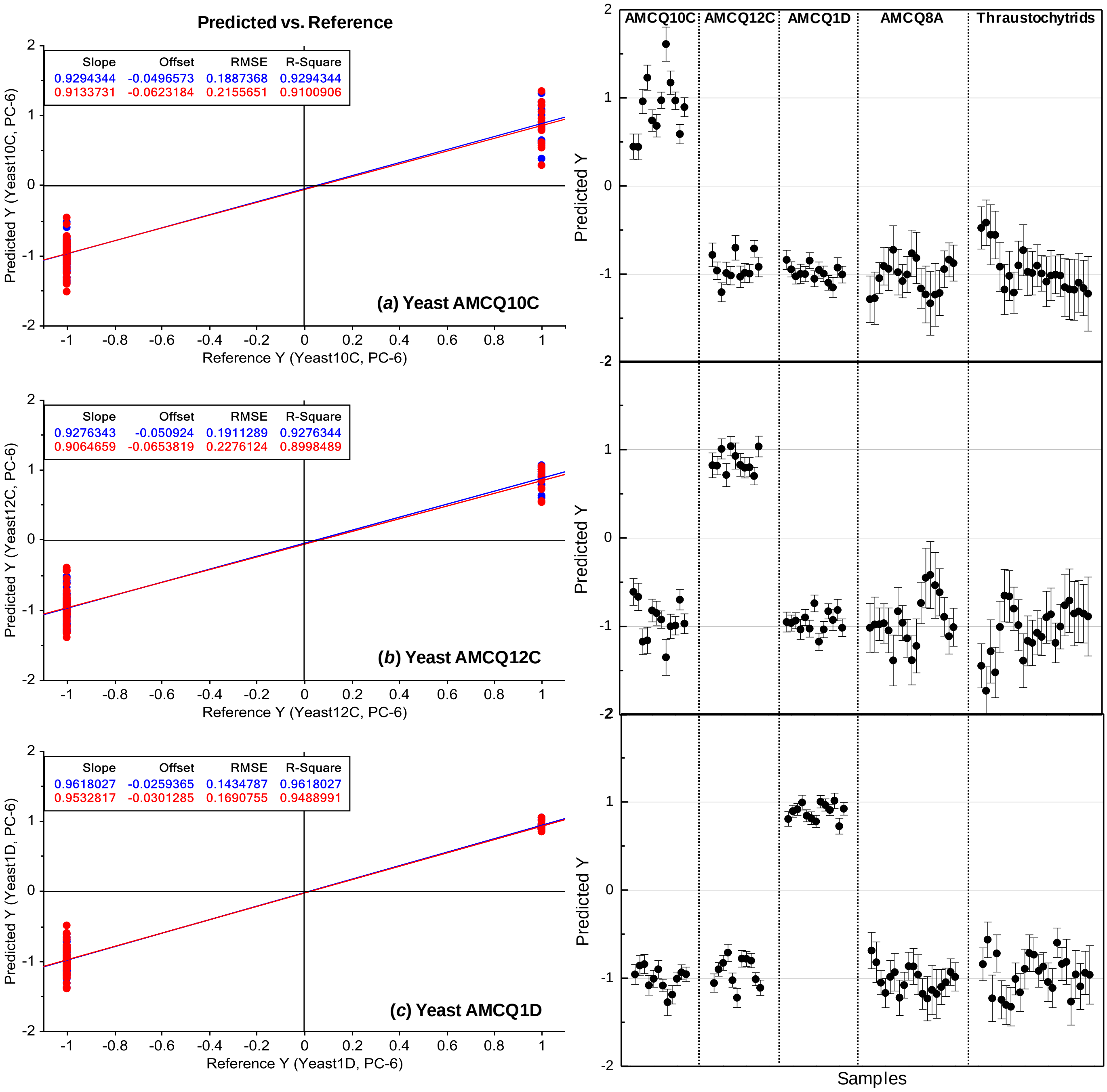


Figure S3: Complementary PLS-DA results to that presented in Fig. 4 including linear regression models trained by replicate II spectral set (*left*) and their corresponding prediction results of replicate I samples used as an independent validation set (*right*). Note that the numbers of the cell samples included in replicate I and II sets are 81 and 79, respectively.



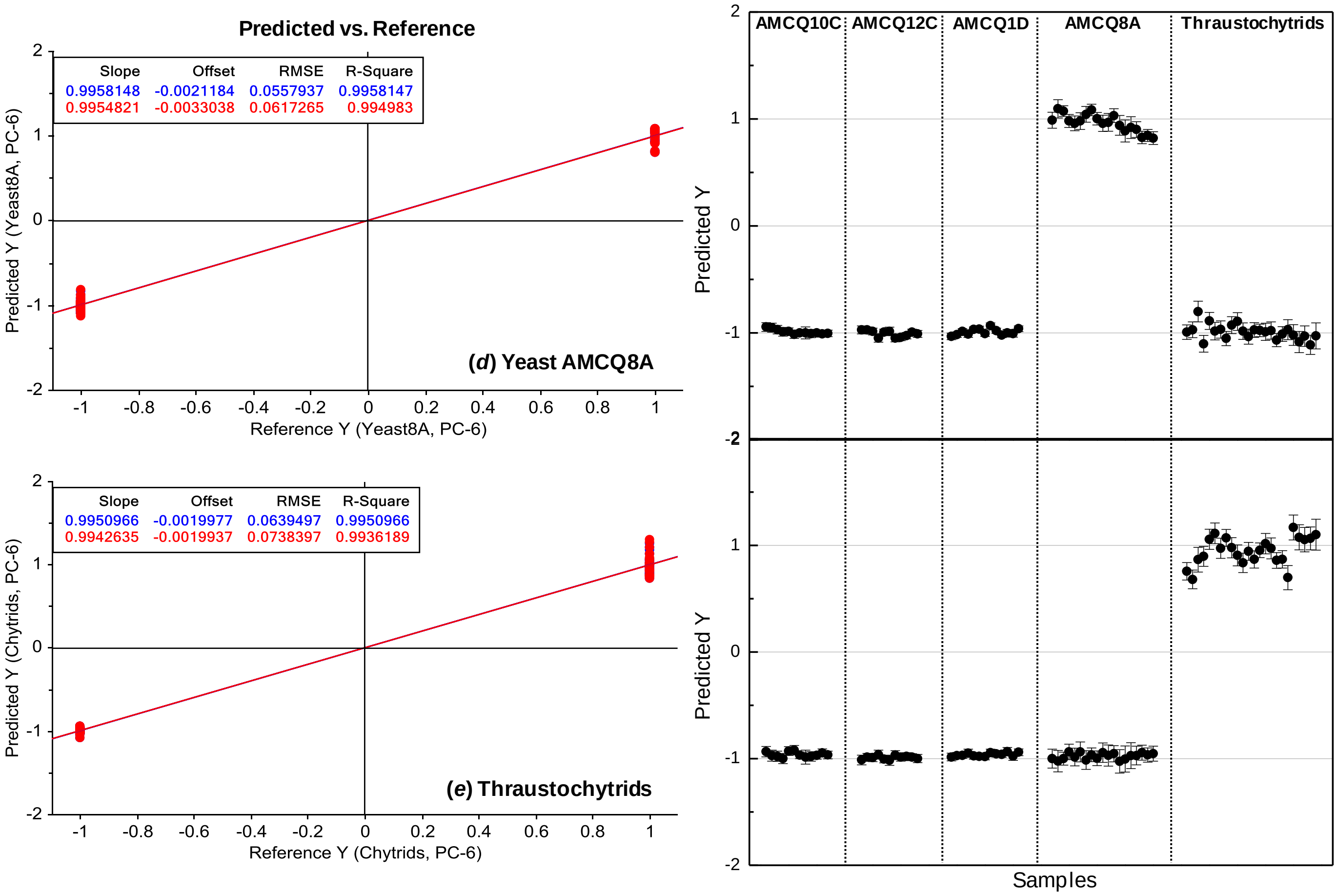
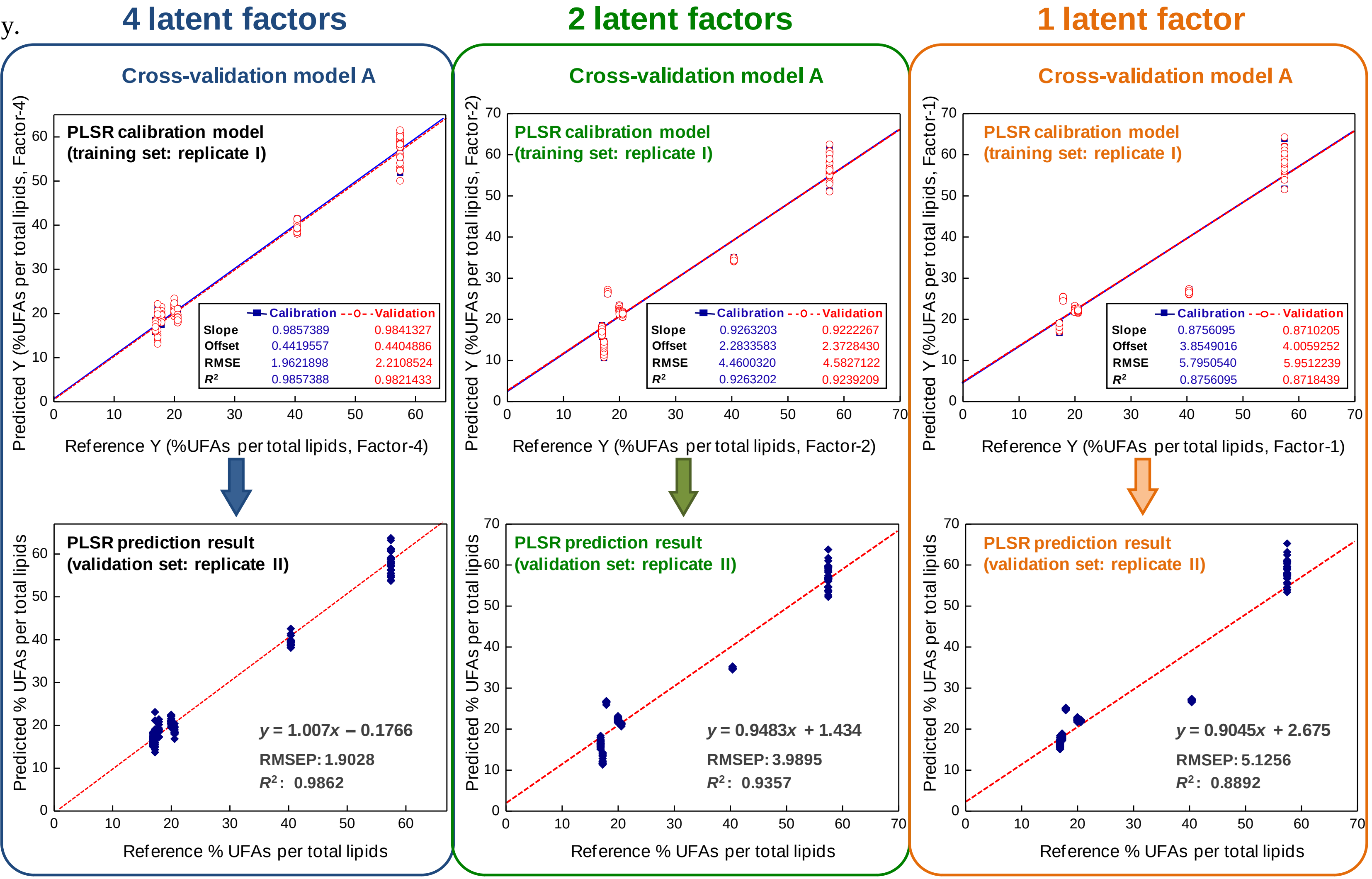

Predicted vs. Reference
Slope
Offset
RMSE
R-Square
0.9958148
-0.0021184
0.0557937
0.9958147
0.9954821
-0.0033038
0.0617265
0.994983
Predicted Y (Yeast8A, PC-6)
Reference Y (Yeast8A, PC-6)
(d) Yeast AMCQ8A
Slope
Offset
RMSE
R-Square
0.9950966
-0.0019977
0.0639497
0.9950966
0.9942635
-0.0019937
0.0738397
0.9936189
Predicted Y (Chytrids, PC-6)
Reference Y (Chytrids, PC-6)
(e) Thraustochytrids
2
1
0
-1
-2
-1
-0.8
-0.6
-0.4
-0.2
0
0.2
0.4
0.6
0.8
1
AMCQ10C
AMCQ12C
AMCQ1D
AMCQ8A
Thraustochytrids
Predicted Y
Samples

Table S4: Complementary SIMCA classification results at 95% significance limit obtained based on the cross-validation approach using replicate II spectral data as a training set and spectra in replicate I set as independent validation (test) samples, with the same parameters used in Table 2. Note that the numbers of the cell samples included in replicate I and II sets are 81 and 79, respectively.

Samples	**Yeasts *Rhodotorula* sp.**				**Thraustochytrids**
Class membership 5%	**AMCQ10C**	**AMCQ12C**	**AMCQ1D**	**AMCQ8A**	
10C-R1_02	*				
10C-R1_03	*				
10C-R1_05	*				
10C-R1_08	*				
10C-R1_09	*				
10C-R1_11	*				
10C-R1_12	*				
10C-R1_13	*				
10C-R1_17	*				
10C-R1_20	*				
10C-R1_21	*				
12C-R1_02		*			
12C-R1_04	*	*			
12C-R1_06		*			
12C-R1_07		*			
12C-R1_08		*			
12C-R1_11		*			
12C-R1_12		*			
12C-R1_14		*			
12C-R1_15		*			
12C-R1_18		*			
12C-R1_19		*			
12C-R1_20		*			
1D-R1_01			*		
1D-R1_02			*		
1D-R1_04			*		
1D-R1_07			*		
1D-R1_08			*		
1D-R1_10			*		
1D-R1_13			*		
1D-R1_14			*		
1D-R1_17			*		
1D-R1_18			*		
1D-R1_19			*		
1D-R1_20			*		
1D-R1_21			*		

8A-GC5-R1_11	*	
8A-GC5-R1_15	*	
8A-GC5-R1_17	*	
8A-GC5-R1_19	*	
8A-GC5-R1_20	*	
8A-GC5-R1_21	*	
8A-GC5-R1_25	*	
8A-GC5-R1_26	*	
8A-GC5-R1_28	*	
8A-GC5-R1_31	*	
8A-GC5-R1_32	*	
8A-GC5-R1_34	*	
8A-GC5-R1_35	*	
8A-GC5-R1_37	*	
8A-GC5-R1_38	*	
8A-GC5-R1_43	*	
8A-GC5-R1_44	*	
8A-GC5-R1_46	*	
8A-GC5-R1_49	*	
8A-GC5-R1_50	*	
PRA-R1_01		*
PRA-R1_04		*
PRA-R1_05		*
PRA-R1_07		*
PRA-R1_10		*
PRA-R1_11		*
PRA-R1_13		*
PRA-R1_14		*
PRA-R1_16		*
PRA-R1_17		*
PRA-R1_19		*
PRA-R1_20		*
PRA-R1_21		*
PRA-R1_24		*
PRA-R1_25		*
PRA-R1_27		*
PRA-R1_28		*
PRA-R1_30		*
PRA-R1_33		*
PRA-R1_34		*
PRA-R1_36		*
PRA-R1_39		*
PRA-R1_40		*
PRA-R1_42		*

Figure S5: Optimised PLSR calibration models and the predictions for quantitative determination of the time-course %UFAs based on cross-validation approach using 2nd derivative spectral input over the two spectral windows that contain biological information about the cells (i.e. 3100-2800 and 1800-965 cm^{-1}) and different number of latent factors. Note that the numbers of the cell samples included in replicate I and II sets are 81 and 79, respectively.



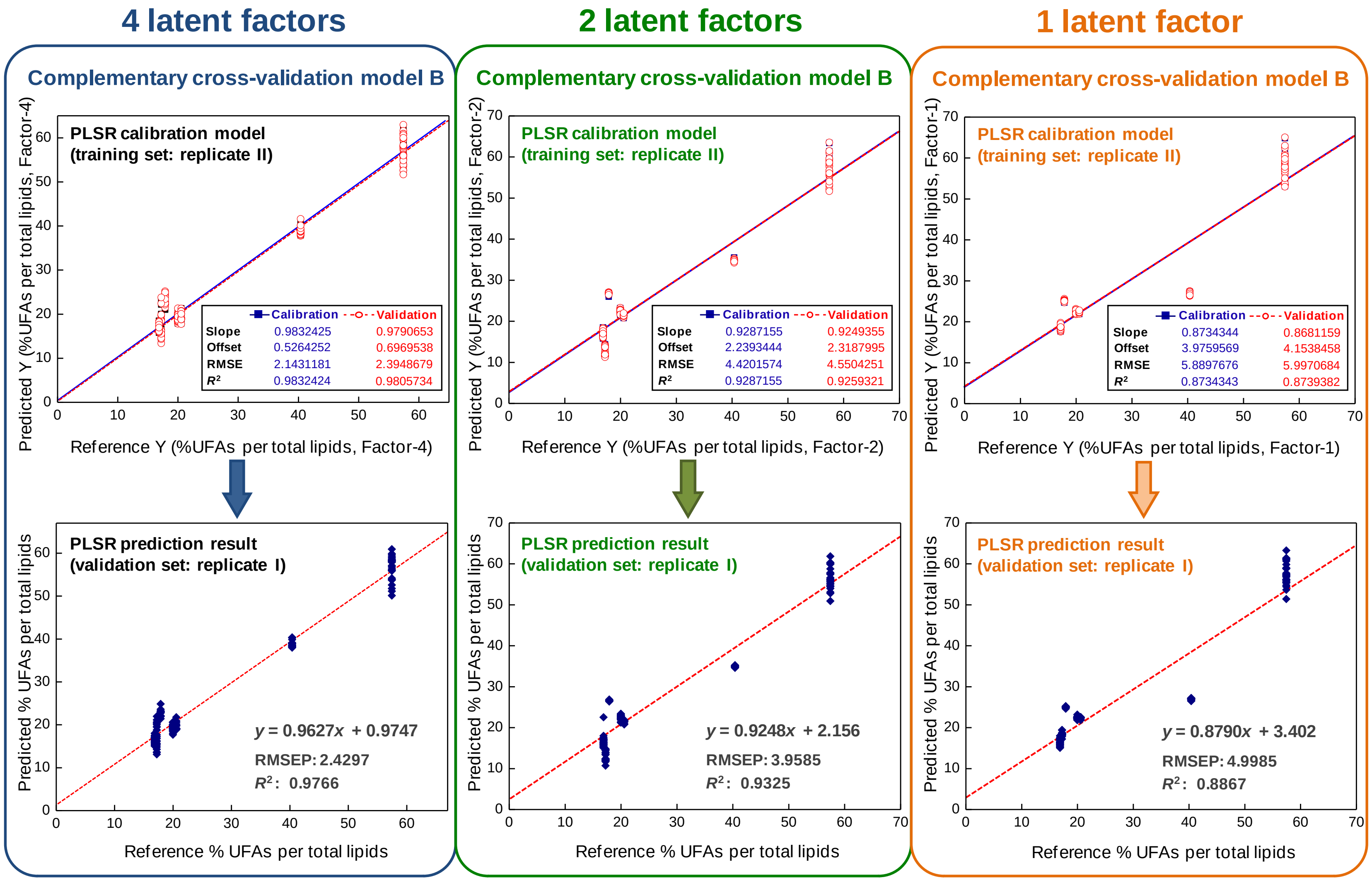

4 latent factors
Complementary cross-validation model B
PLSR calibration model
(training set: replicate II)
Calibration
Validation
Slope 0.9832425 0.9790653
Offset 0.5264252 0.6969538
RMSE 2.1431181 2.3948679
R² 0.9832424 0.9805734
Predicted Y (%UFAs per total lipids, Factor-4)
Reference Y (%UFAs per total lipids, Factor-4)
PLSR prediction result
(validation set: replicate I)
y = 0.9627x + 0.9747
RMSEP: 2.4297
R²: 0.9766
Predicted % UFAs per total lipids
Reference % UFAs per total lipids
2 latent factors
Complementary cross-validation model B
PLSR calibration model
(training set: replicate II)
Calibration
Validation
Slope 0.9287155 0.9249355
Offset 2.2393444 2.3187995
RMSE 4.4201574 4.5504251
R² 0.9287155 0.9259321
Predicted Y (%UFAs per total lipids, Factor-2)
Reference Y (%UFAs per total lipids, Factor-2)
PLSR prediction result
(validation set: replicate I)
y = 0.9248x + 2.156
RMSEP: 3.9585
R²: 0.9325
Predicted % UFAs per total lipids
Reference % UFAs per total lipids
1 latent factor
Complementary cross-validation model B
PLSR calibration model
(training set: replicate II)
Calibration
Validation
Slope 0.8734344 0.8681159
Offset 3.9759569 4.1538458
RMSE 5.8897676 5.9970684
R² 0.8734343 0.8739382
Predicted Y (%UFAs per total lipids, Factor-1)
Reference Y (%UFAs per total lipids, Factor-1)
PLSR prediction result
(validation set: replicate I)
y = 0.8790x + 3.402
RMSEP: 4.9985
R²: 0.8867
Predicted % UFAs per total lipids
Reference % UFAs per total lipids

Figure S6: Comparison of the corresponding PLSR regression coefficients obtained from the optimised calibration models with different number of latent factors as shown in Fig. S5, relative to 2nd derivative spectra of the yeast and the thraustochytrium cells, in order to confirm that the optimised PLSR models and the predictions genuinely reflect spectral information without the contribution from the spectral noise.