Electronic Supplementary Material (ESI) for Physical Chemistry Chemical Physics.
This journal is © the Owner Societies 2017

SUPPORTING INFORMATION

Changes in dynamics of α-chymotrypsin due to covalent inhibitors investigated by elastic incoherent neutron scattering

C. D. Andersson,[a] N. Martinez,[b,c] D. Zeller,[b,c] S. H. Rondahl,[d] M. M. Koza,[b] B. Frick,[b] F. Ekström,[d] J. Peters,[b,c,*] and A. Linusson[a,*]

[a] Department of Chemistry, Umeå University, SE-90187 Umeå, Sweden.
[b] Institut Laue Langevin, F-38042 Grenoble Cedex 9, France.
[c] Univ. Grenoble Alpes, IBS and LiPhy, F-38044 Grenoble, France.
[d] CBRN Defence and Security, Swedish Defence Research Agency, SE-90621 Umeå, Sweden

Contents

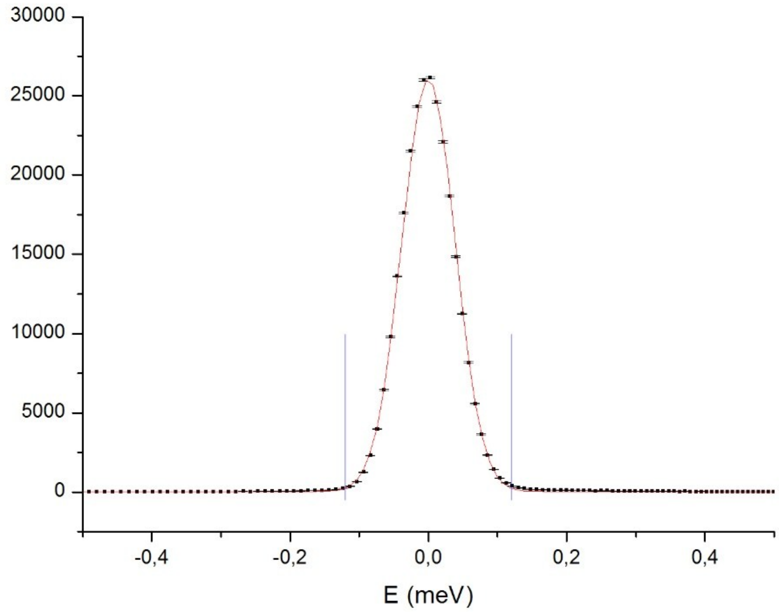

Figure S1. Experimental resolution function (obtained from data of CT2 at 20 K) of IN6 fitted with a Gaussian function (red line). The extracted full width at half maximum is ≈ 91 µeV and the corresponding resolution time window ≈ 7.5 ps.[1] The vertical blue lines indicate the limits taken to integrate the elastic peak.

Multivariate Modelling

Principal Component Analysis

In the data matrix used in PCA, the row elements are called observations and column elements are variables. The EINS data consisted of intensities where the observations are the measurements in sequential order (at different temperatures) and the variables are the Q:s. For this matrix, the variables were centered (subtraction of the mean for each variable). No scaling was performed since the variables are within a similar numerical range. PCA is an unsupervised projection method that transforms an original multidimensional set of correlated variables to a smaller set of uncorrelated variable, so called principal components (PCs) that contain the main variation in the data. These PCs are eigenvectors that approximate the data as well as possible in the least squares sense. The first component PC1 contains the main variation in the data, PC2 the second largest variation and so on. The nonlinear iterative partial least square (NIPALS) algorithm [2] was used to extract the new eigenvectors. The calculation of PCs, score values (*s*) and loadings values (*l*) are exemplified in figure S2.

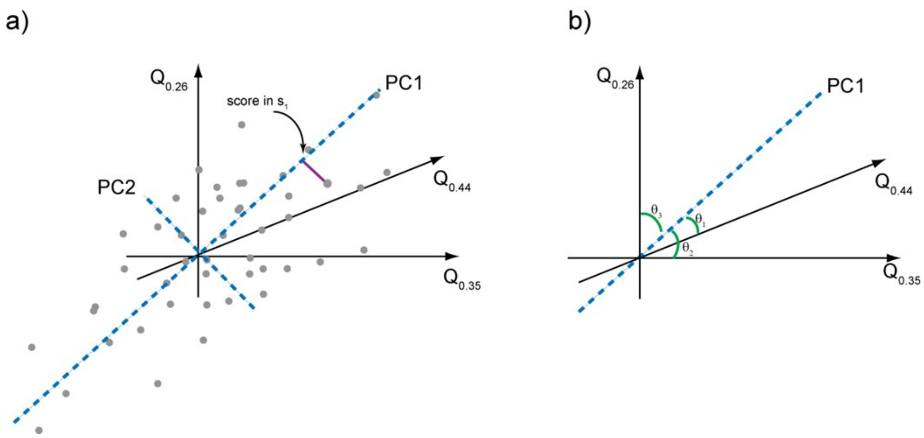

Figure S2. a) A contrived example of the geometrical properties of two principal component (PC) from 46 measurements of three Q:s (0.26, 0.35 and 0.44 Å^{-1}) indicated with gray dots. The purple line represents the projection of a data point on to PC1 yielding a score value for that measurement. b) Loading values for each of the three variables ($Q_{0.44}$, $Q_{0.35}$, $Q_{0.26}$) for PC1 is calculated by the cosine of the corresponding angle (θ_1, θ_2, θ_3) between the original variable and the score PC1. The same is done for PC2, PC3, etc.

PC1 (s_1) is per definition placed in the direction of the main variation in the data (Figure S2a). PC2 is placed orthogonally to PC1, *and* in the direction of the main variation in the remaining data. Each data point is projected down onto PC1 yielding a new variable value, *i.e.* a score value in PC1. The same is done for PC2, PC3, etc. The direction of PC1 is given by the cosine of the angles θ_1, θ_2, θ_3 of the corresponding three original variables ($Q_{0.44}$, $Q_{0.35}$, and $Q_{0.26}$) giving rise to a weight, indicating how much each variable contributes to the direction of the PC (Figure S2b). The same is done for PC2, PC3, etc. The PCs forms a new decomposition matrix according to:

$$X = \bar{X} + SL' + E = \bar{X} + s_1 l_1' + s_2 l_2' + \dots + s_A l_A' + E \quad \text{(S1)}$$

where $\bar{X}$ is the **X** matrix average, **S** is the score vector matrix, **L**´ is the transposed loading vector matrix, **E** is the residual, **s** is a score vector, **l**´ is a loading vector and *A* is the number of extracted PCs. The score values and loading values are visualized in two separate plots. PC1 will show the largest variation in the data. The score plot will show how the observations (EINS measurements) relate to one another; observations with similar variable values (intensities in the different Q:s) will be close in the plot and those with very different variable values will be far apart. The loading plot shows how much each original variable contributes to each PC; the further away from the origin (large absolute values) in the direction of the PC axis, the more a variable contributes. High contribution of a variable means that that variable varies a lot between the measurements. Interpretation of the score- and loading plot together gives information on the most similar and dissimilar observations and the variables that contain data giving rise to these similarities and differences.

OPLS-DA

The orthogonal partial least square regression discriminant analysis (OPLS-DA) has its mathematical basis in orthogonal signal correction (OSC) and partial least square regression (PLS) giving the method OPLS [3]. PLS is a linear regression method and an extension of PCA which, instead of maximizing the explained variance in **X**, maximizes the covariance between **X** and a second matrix (or vector) **Y**. In OPLS, OSC is used to maximize the explained covariation between **X** and **Y** on the first OPLS component and the second (and higher) components capture variance in **X** that is not related to **Y**. This regression method can be extended to classification by OPLS discriminant analysis (OPLS-DA)[4] where the **Y** contains *pre-defined* class memberships and **X** the observations and variables (Figure S3). OPLS-DA separates the discriminatory direction (prediction of class) in component on ($\mathbf{s}_{p,1}$) from the Y-orthogonal direction ($\mathbf{s}_{o,1}$), making the predictive weight (loading) vector $\mathbf{w}_{p,1}$ more straightforward to interpret.

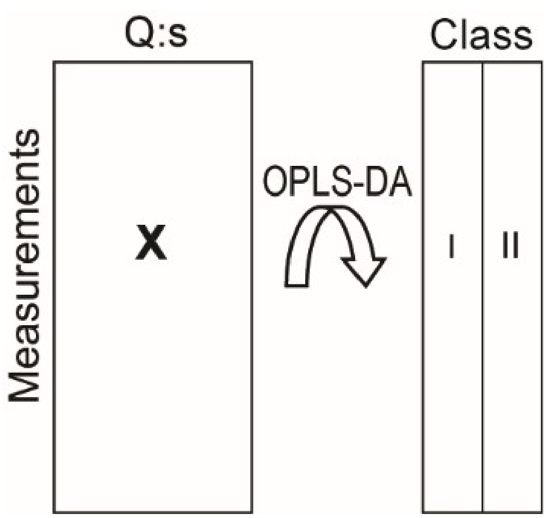

Figure S3. Explanation of OPLS-DA matrices, where **X** is the data matrix consisting of the EINS data and **Y** is the matrix containing class memberships.

Cross Validation

The general procedure for cross validation is as follows. 1) 1/7 of the **X**-data are kept out of model calculation and 2) the excluded data are predicted by the model calculated on the remaining X-data. 3) The predictions are compared with the actual values. 4) The procedure is repeated until all parts have been kept out once [5]. The prediction error sum of squares (PRESS) describes the difference between the predicted and true values for the **X**-data kept out of the model fitting.

In PCA, the prediction of the (i, k) element in centered and scaled form is the i^{th} score value (*s*) multiplied by the k^{th} loading value (*l*), where both have been estimated in a cross validation round when this element was excluded. For each principal component consecutively, the overall PRESS/SS is computed, where SS is the residual sum of squares of the previous component. The total cross validated $q^2(cum)=1-\Sigma(PRESS/SS)_a$, where $\Sigma(PRESS/SS)_a$ the product of PRESS/SS for each individual principal component *a.*

The OPLS cross-validation is performed as follows. *Step1*: The number of predictive or joint **X**/**Y** components is estimated; this number may be adjusted using step2 after calculating all predictive and orthogonal components. *Step2*: The number of orthogonal components in X and Y are determined. For these steps, step1 is used to determine the significance of the components. For all included components, the $q^2=1-PRESS/SS$ is computed, where SS is the sum of squares of **Y**. The cross validation for the orthogonal in **X** (PCA) and the orthogonal in **Y** (PCA) is performed as for the regular PCA described above. The total cross validated $q^2(cum)=1-\Sigma(PRESS/SS)_a$, where $\Sigma(PRESS/SS)_a$ the product of PRESS/SS for each individual OPLS component *a.*

Statistical evaluation of PCA and OPLS-DA models.

Table S1. IN6 model statistics.

Model	Samples (class)	Num. of Measure-ments, *N*	Num. of Variables, *M*	Temp. (K)[a]	Num. of Comp-onents	PC1, R^2**X** (Eig.)	PC2, R^2**X** (Eig.)	PC3, R^2**X** (Eig)	R^2**X** Cumul.	R^2**Y** Cumul.	q^2 cumul.
PCA	CT1	20	28	80-310	3	0.967 (27.1)	0.01 (0.4)	0.007 (0.2)	0.988	-	0.980
	CT2	21									
	CT/CS	20									
	CT/TP	20									
OPLS-DA	CT1(1)	4	28	275-310	2				0.966	0.976	0.923
	CT2(1,	4									
	CT/CS(2)	4									
	CT/TP(2)	4									

[a] actual temperature in the measurement deviated slightly from the given range.

Table S2. IN13 model statistics.

Model	Samples (class)	Num. of Measure-ments, *N*	Num. of Variables, *M*	Temp. (K)[a]	Num. of Comp-onents	PC1, R^2**X** (Eig.)	PC2, R^2**X** (Eig.)	R^2**X** Cumul.	R^2**Y** Cumul.	q^2 Cumul.
PCA	CT2	27	22	40-310	2	0.972 (21.4)	0.016 (0.3)	0.988	-	0.985
	CT/CS	27								
	CT/TP	26								
OPLS-DA	CT2(1)	5	22	270-315	2			0.954	0.902	0.857
	CT/CS(2)	5								
	CT/TP(2)	5								

[a] actual temperature in the measurement deviated slightly from the given range.

Table S3. IN16B model statistics

Model	Samples (class)	Num. of Measure-ments, *N*	Num. of Variables, *M*	Temp. (K)[a]	Num. of Comp-onents	PC1, R^2**X** (Eig.)	PC2, R^2**X** (Eig.)	R^2**X** Cumul.	R^2**Y** Cumul.	q^2 Cumul.
PCA	CT2	49	14	40-310	2	0.995 (13.9)	0.0045 (0.06)	0.999	-	0.999
	CT/CS	48								
	CT/TP	49								
OPLS-DA	CT2(1)	8	14	270-310	2			0.999	0.99	0.988
	CT/CS(2)	8								
	CT/TP(2)	8								

[a] actual temperature in the measurement deviated slightly from the given range.

OPLS-DA of IN6 measurement

OPLS-DA model of the IN6 data (Figure S4) including the whole temperature range (80-310 K). Class one constituted the CT1 and CT2 samples and class 2 the CT/CS and CT/TP samples. The model had two OPLS-components with an R^2**X**(cum) of 0.98, R^2**Y**(cum) of 0.95, and q^2(cum) of 0.94.

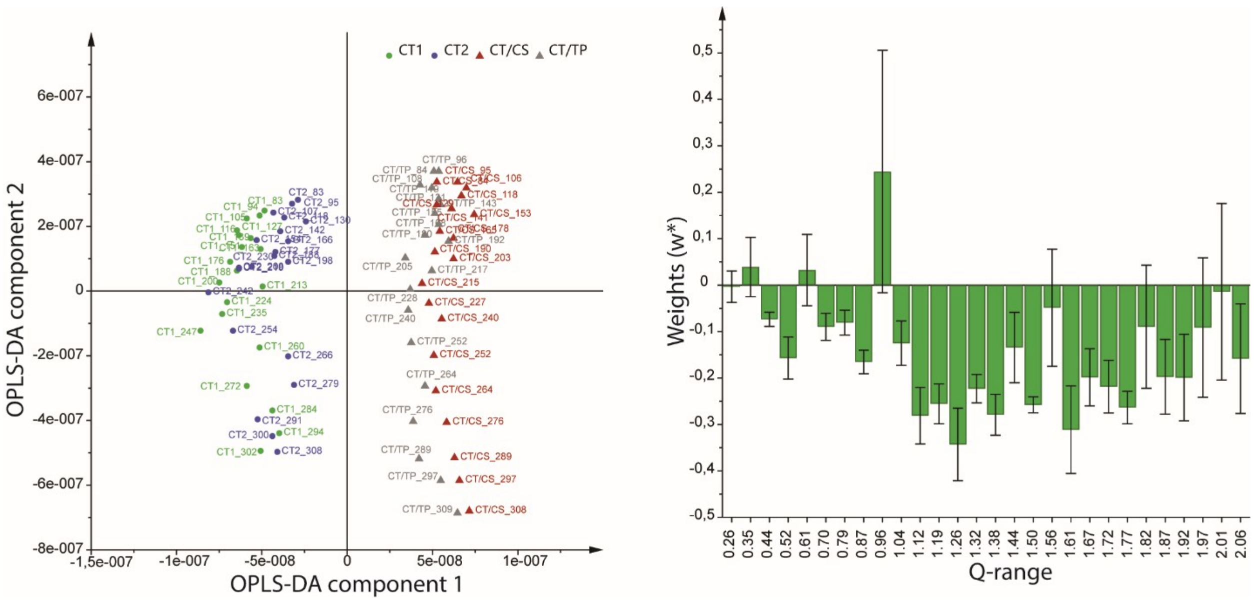

Figure S4. OPLS-DA model from the IN6 measurements including a) the score plot and b) the weights plot.

References

1. Magazù, S.; Migliardo, F.; Benedetto, A., *Rev. Sci. Instrum.*, 2011, **82**, 105115.
2. Wold, H., *Nonlinear estimation by iterative least square procedures*. Wiley: New York, USA, 1966; p 411-444.
3. Trygg, J.; Wold, S., *J. Chemometr.*, 2002, **16**, 119-128.
4. Bylesjö, M.; Rantalainen, M.; Cloarec, O.; Nicholson, J. K.; Holmes, E.; Trygg, J., *J. Chemometr.*, 2006, **20**, 341-351.
5. Eastment, H. T.; Krzanowski, W. J., *Technometrics*, 1982, **24**, 73-77.