DOI: 10.1002/cmdc.200700085

SERS Classification of Highly Related Performance Enhancers

Ramon A. Alvarez-Puebla,[a] Juan P. Bravo-Vasquez,[a] Bo Cui,[b] Teodor Veres,[b] and Hicham Fenniri*[a, c]

Detection of doping represents a top priority for sports organizations and athletic governing bodies because of the illegal enhancement of athletic performance and also the potential health risks to the athletes.[1] Recombinant protein abuse is an evolving form of doping that challenges analytical detection. The World Anti-Doping Agency (WADA)[2] prohibits the use of recombinant proteins that are classified under doping substances. These substances are difficult to detect because they are very similar to their endogenous counterparts. Corticotropin releasing factors (CRFs)[3] are polypeptides produced in the hypothalamus that induce the secretion of corticotrophin (ACTH) and β-endorphins from the anterior pituitary. This communication demonstrates a) the identification of three closely related CRFs by surface-enhanced Raman scattering (SERS) spectroscopy, b) their differentiation with 100% confidence using principal component regression (PCR) analysis, and c) the detection of very small amounts (<10%) of nonhuman CRF co-mixed with human samples by performing a prediction based on the spectral data.

Consistent with their primary sequences (see Experimental Section), human (H), bovine (B), and sheep (S) CRFs show very similar vibrational patterns in their Raman spectra (Figure 1A). The spectra are characterized by strong peaks at 1654 cm^{-1} (amide I, α-helix),[4] 1448 cm^{-1} (H–C–H scissoring), 1342 cm^{-1} (H–C–H deformation), 1002 cm^{-1} (ring breathing) and 933 cm^{-1} (C–C skeletal stretching).[5–7] Although Raman spectroscopy is a powerful analytical tool for structural characterization,[8,9] its low cross-section (10^{-29} cm^2 molecule^{-1}) limits its application in biodiagnostics. In contrast with this, the molecular cross-section increases significantly in SERS spectroscopy (10^{-17} cm^2 molecule^{-1}), allowing ultrasensitive chemical analysis and single-molecule spectroscopy of peptides and proteins.[10–12] For SERS analysis, the analyte must be adsorbed onto an optical enhancer, usually nanostructured silver or gold. For this we designed a nano-imprinted substrate upon which a 9-nm Ag island film was evaporated with a maximum surface plasmon resonance (SPR) at 710 nm (Figure 1B).[13] CRF films,

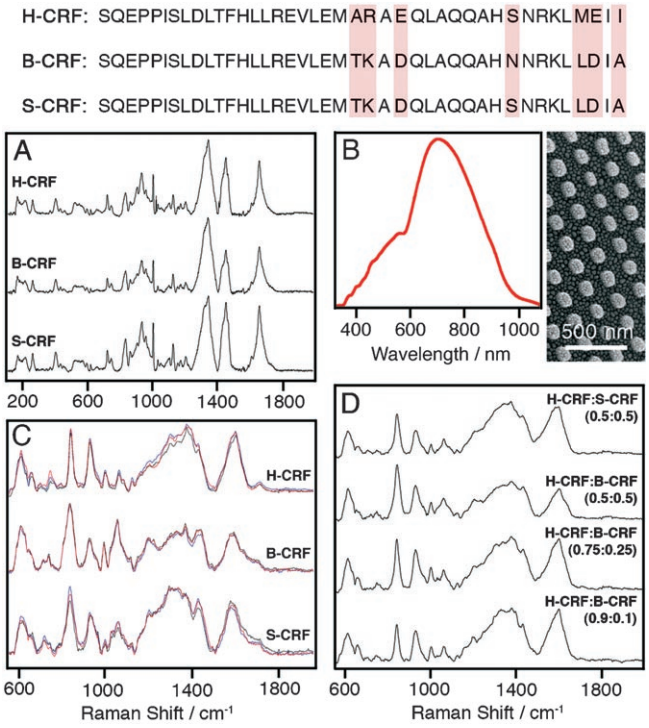

Figure 1. Raman and SERS spectra of human, bovine, and sheep CRFs on a nanostructured optical enhancer. A) Raman spectra of H-CRF, B-CRF, S-CRFs. B) Surface plasmon resonance spectrum and SEM image of the pillared substrate used as optical enhancer. C) Triplicate SERS spectra of the pure CRFs. D) Average SERS spectra (10 points co-added) of CRF mixtures.

prepared by casting solutions (10^{-7} M) and air-drying, were studied with the 780-nm laser line. To test for reproducibility, a mapping tool was used, collecting 266 spectra per CRF in triplicate (798 spectra per sample). The SERS spectra of the CRFs (Figure 1C) and their mixtures (Figure 1D) show similar vibrational features, but with some characteristic peak differences. In agreement with surface selection rules,[14] these differences are likely due to subtle conformational variations among the CRFs,[15] which are amplified in SERS.

Multivariate data analysis was used to compare and classify the SERS spectra of CRFs. First, two matrices were generated: 1) a reference identity matrix (30×3) that defines the composition of 30 samples (10 samples for each type of CRF) and three variables (H-CRF, B-CRF, S-CRF), and 2) a spectral matrix (30×879) that comprises 30 samples (10 samples for each CRF) and 879 variables per SERS spectrum. Then, a partial least-squares regression (PLS2) model that correlates these two matrices was generated. Figure 2A shows the score plot for the first and second principal components (PCs) of the regression model. The first two PCs accounted for 95% of the variance in the identity matrix. Notably, the CRFs were clustered based on their origin (gray regions, Figure 2A), thus demonstrating that closely related CRFs could be classified with high confidence based on their SERS spectra.

The validity of the regression model was established by generating a predicted identity matrix (*Y'*) from the spectral matrix (*X*), and correlating it to the reference identity matrix (*Y*). The correlation between *Y* and *Y'* was obtained from the inner

[a] *Dr. R. A. Alvarez-Puebla, Dr. J. P. Bravo-Vasquez, Prof. Dr. H. Fenniri*
National Institute for Nanotechnology (NINT), National Research Council
11421 Saskatchewan Drive, Edmonton, AB, T6G 2M9 (Canada)
E-mail: hicham.fenniri@ualberta.ca

[b] *Dr. B. Cui, Dr. T. Veres*
Industrial Materials Institute, National Research Council
75 Boulevard de Mortagne, Boucherville, QC, J4B 6Y4 (Canada)

[c] *Prof. Dr. H. Fenniri*
Department of Chemistry, University of Alberta
Edmonton, AB, T6G 2G2 (Canada)
Fax: (+1) 780-641-1601

Supporting information for this article is available on the WWW under http://www.chemmedchem.org or from the author.

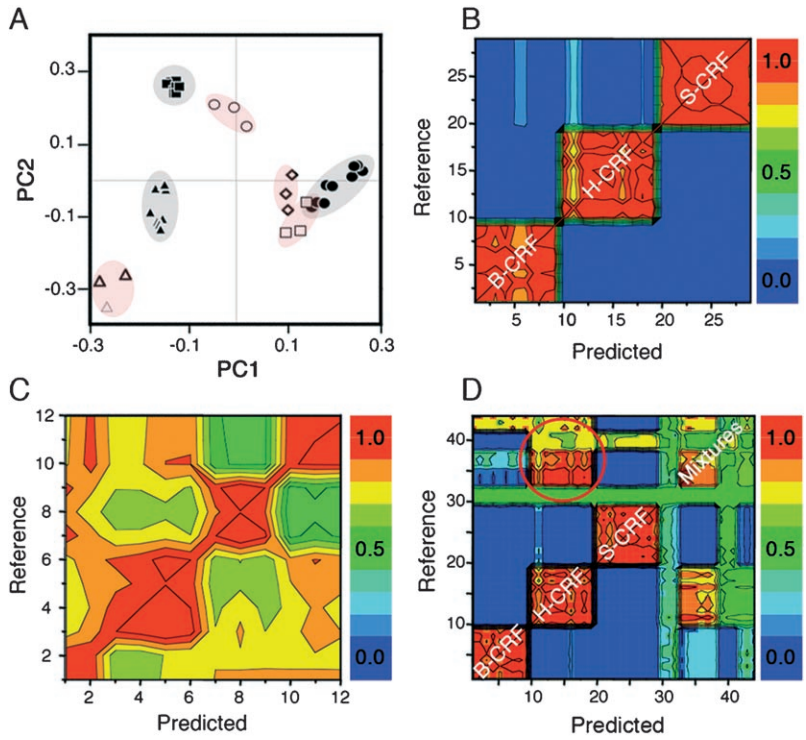

Figure 2. Classification of human, bovine, and sheep CRFs, and mixtures thereof. *X* and *Y* axes refer to the predicted and reference samples, respectively. A) Score plot representing the first and second PCs for the PLS2 regression model of human (●, H-CRF), bovine (■, B-CRF) and sheep (▲, S-CRF) CRFs, and mixtures thereof (□, 0.9:0.1 H-CRF/B-CRF; ◇, 0.75:0.25 H-CRF/B-CRF; △, 0.5:0.5 H-CRF/S-CRF; ○, 0.5:0.5 H-CRF/B-CRF). B) Contour plot showing the angular correlation between the reference and predicted identity matrices for B-CRF (samples 1–9), H-CRF (samples 10–19) and S-CRF (samples 20–29). C) Contour plot showing the angular correlation between the reference and predicted identity matrices for H-CRF/B-CRF/S-CRF 1:1:1 (samples 1, 2), H-CRF/B-CRF 0.9:0.1 (samples 3, 4), H-CRF/B-CRF 0.75:0.25 (samples 5, 6), H-CRF/S-CRF 0.5:0.5 (samples 7–9), and H-CRF/B-CRF 0.5:0.5 (samples 10–12). D) Contour plot showing the angular correlation between the reference and predicted identity matrices for B-CRF (samples 1–9), H-CRF (samples 10–19), S-CRF (samples 20–29), H-CRF/B-CRF/S-CRF 1:1:1 (samples 30, 31), H-CRF/B-CRF 0.9:0.1 (samples 32, 33), H-CRF/B-CRF 0.75:0.25 (samples 34, 35), H-CRF/S-CRF 0.5:0.5 (samples 36–38), and H-CRF/B-CRF 0.5:0.5 (samples 39–41).

product [Eq. (1)] of the reference composition vectors y_i and the predicted identity vectors y'_i unique to each sample. The angular matrix resulting from the inner product of *Y* and *Y'* is shown as a contour plot (Figure 2B). Cos θ values of ~1 (highest correlation) were obtained for identity vectors corresponding to CRFs of the same origin, whereas no correlation was found between samples of different origin (blue regions in Figure 2B).

$$\cos\theta = \frac{y_i y'_i}{|y_i||y'_i|} \qquad (1)$$

To test the feasibility of detecting and determining the level of contamination in unknown samples, a PLS2 regression model was generated for the CRF mixtures in the same manner as for the pure CRF samples (see above). The first two PCs of the regression model accounted for 86% of the variance in the identity matrix and resulted in a clear distinction between the CRF mixtures (pink regions, Figure 2A). The inner product matrix represented as a contour plot in Figure 2C shows the highest level of correlation between reference and predicted identity vectors corresponding to CRF mixtures with the same composition (red diagonal). However, the model also produces strong off-diagonal correlations resulting from the presence of exogenous CRFs in the human samples. This result establishes that contaminated samples could be readily identified and to some extent quantified.

To further demonstrate the relative ease with which a contaminated sample could be classified, a contour plot featuring the angular correlation between all the samples (Figure 2D) was generated as described above for the pure (Figure 2B) and mixture samples (Figure 2C). Once again, the pure and mixture samples showed excellent correlation between the relevant predicted and reference identity vectors (normalized diagonal values for cos θ exceed 0.7). However, the appearance of off-diagonal angular values >0.6 and as high as 1.0, demonstrates the ability of the regression model to identify contaminated samples (mixture samples) relative to pure samples. For instance, the region within the red circle in Figure 2D shows that all the mixture samples contain H-CRF, and as the percentage of H-CRF decreases in the sample, the correlation also decreases.

In summary, we have shown that peptidic performance enhancers with more than 80% sequence homology could be classified with 100% confidence from simple SERS measurements and multivariate data analysis. We have also shown that 'contaminated' samples (mixtures) can be readily identified, classified, and even quantified by using the same approach. This method may constitute a fast and reliable approach for the rapid identification of doping peptides, and may be extended to other important proteogenic enhancers such as erythropoietin and growth hormones.

Experimental Section

Materials: Dilute solutions (10^{-7} M) of the pure CRFs (Sigma–Aldrich) were prepared using phosphate-buffered saline (PBS, 0.05 M, pH 7.4). Mixtures of CRFs were prepared by mixing the appropriate volumes from the CRF stock solutions.

Raman microspectroscopy: Raman and SERS spectra were collected on a Nicolet Almega system using a near-IR excitation laser line (780 nm) to minimize photobleaching. Raman spectra were collect-

ed on 10 different samples for each CRF in high-resolution mode (2400 g mm^{-1} grating). The laser beam was focused onto the sample using a 100× objective, providing a scattering area of ~1 μm^2. Two spectra with accumulation times of 5 s were co-added for each experiment (Figure 1 A). Thin films for SERS were prepared by casting and air-drying 10 μL of a 10^{-7} M solution of each CRF on a nano-imprinted pillared silver island substrate (Figure 1 B), resulting in ~10^4 molecules μm^{-2}. The films were studied using the mapping mode (19×14 points) with a step size of 5 μm. The other acquisition parameters are the same as above. To ensure reproducibility, three different films were mapped, giving rise to a dataset of 798 spectra per sample (Figure 1 C and D).

Multivariate data analysis: The SERS spectra were analyzed and correlated by partial least-squares regression (PLS2) using the Unscrambler software package: a) Pre-processing of all the SERS spectra involved baseline correction, normalization and Norris-Gap first derivative. b) The reference identity matrix (*Y*) composed of the samples' identities (human, bovine, sheep, or mixtures thereof) was generated. c) The spectral matrix (*X*) was generated from the SERS spectra (879 variables per spectrum) for each sample (human, bovine, sheep, or mixtures thereof). d) A regression model that correlates the identity and spectral matrices was generated using a PLS2 regression. e) The validity of the regression model was established by generating a predicted identity matrix (*Y'*) from the spectral matrix (*X*), and correlating it to the reference identity matrix (*Y*). The correlation between *Y* and *Y'* was obtained from the inner product [Eq. (1)] of the reference composition vectors y_i and the predicted identity vectors y'_i unique to each sample. If y_i and y'_i are highly correlated (that is, $y_i = y'_i$), $\cos\theta \sim 1$. f) The angular matrix resulting from the inner product of *Y* and *Y'* is better visualized in a contour plot (Figure 2 B). The same sequence of steps (e, f) was used to compare the mixture samples (Figure 2 C and 2D).

Acknowledgements

We thank the National Research Council (NRC) of Canada, the Genomics and Health Initiative, the National Institute for Nanotechnology, the Industrial Materials Institute, and the University of Alberta.

Keywords: corticotrophin releasing factor · doping drugs · multivariate data analysis · SERS · sports

[1] G. Vogel, *Science* **2004**, *305*, 632 – 635.
[2] http://www.wada-ama.org.
[3] W. Vale, J. Spiess, C. Rivier, J. Rivier, *Science* **1981**, *213*, 1394 – 1397.
[4] M. E. Rousseau, T. Lefevre, L. Beaulieu, T. Asakura, M. Pezolet, *Biomacromolecules* **2004**, *5*, 2247 – 2257.
[5] W. L. Peticolas, *Methods Enzymol.* **1995**, 389.
[6] J. E. LeBihan, A. M. Blochet, D. Desormeaux, M. Pezolet, *Biochemistry* **1996**, *35*, 12712 – 12722.
[7] I. Chourpa, V. Ducel, J. Richard, P. Dubois, F. Boury, *Biomacromolecules* **2006**, *7*, 2616 – 2623.
[8] T. E. Keyes, D. Leane, R. J. Forster, N. Moran, D. Kenny, *Proc. SPIE-Int. Soc. Opt. Eng.* **2005**, *5826*, 221 – 231.
[9] R. Tuma, *J. Raman Spectrosc.* **2005**, *36*, 307 – 319.
[10] K. Kneipp, H. Kneipp, S. Abdali, R. W. Berg, H. Bohr, *Spectroscopy* **2004**, *18*, 433 – 440.
[11] S. Habuchi, M. Cotlet, R. Gronheid, G. Dirix, J. Michiels, J. Vanderleyden, F. C. De Schryver, J. Hofkens, *J. Am. Chem. Soc.* **2003**, *125*, 8446 – 8447.
[12] N. Sundararajan, D. Mao, S. Chan, T.-W. Koo, X. Su, L. Sun, J. Zhang, K.-b. Sung, M. Yamakawa, P. R. Gafken, T. Randolph, D. McLerran, Z. Feng, A. A. Berlin, M. B. Roth, *Anal. Chem.* **2006**, *78*, 3543 – 3550.
[13] R. Alvarez-Puebla, B. Cui, J. P. Bravo-Vasquez, T. Veres, H. Fenniri, *J. Phys. Chem. C* **2007**, *111*, 6720 – 6723.
[14] M. Moskovits, *Rev. Mod. Phys.* **1985**, *57*, 783 – 826.
[15] E. Vass, M. Hollosi, F. Besson, R. Buchet, *Chem. Rev.* **2003**, *103*, 1917 – 1954.

Received: April 13, 2007
Published online on May 31, 2007

