New Frontiers in Screening for Microbial Biocatalysts
Edited by K. Kieslich, C.P. van der Beek, J.A.M. de Bont and W.J.J. van den Tweel

DRASTIC (DIFFUSE REFLECTANCE ABSORBANCE SPECTROSCOPY TAKING IN CHEMOMETRICS). A novel, rapid, hyperspectral, FT-IR-based approach to screening for biocatalytic activity and metabolite overproduction

Douglas B. Kell, Michael K. Winson, Royston Goodacre[†], Andrew M. Woodward, Bjørn K. Alsberg, Alun Jones, Éadaoin M. Timmins and [†]Jem J. Rowland

Institute of Biological Sciences and [†]Dept of Computer Science, University of Wales, Aberystwyth, Dyfed SY23 3DA, U.K.
Email: dbk@aber.ac.uk WWW: http://gepasi.dbs.aber.ac.uk/home.htm

Diffuse-reflectance absorbance spectroscopy in the mid-infrared is a novel method of producing data with which to effect chemical imaging for the rapid screening of biological samples for metabolite overproduction. We have used mixtures of ampicillin and *Escherichia coli*, and *Streptomyces citricolor* producing aristeromycin and neplanocin A, as model systems. Deconvolution of the hyperspectral information provided by the raw diffuse reflectance-absorbance mid-infrared spectra may be achieved using a combination of principal components analysis (PCA) and supervised methods such as artificial neural networks (ANNs) and partial least squares regression (PLS). Whereas a univariate approach necessitates appropriate data selection to remove any interferences, the chemometrics/hyperspectral approach could be employed to permit filtering of undesired components either manually, or by taking the Fourier transform of the spectral information (in order to help isolate the signal from the baseline variation or noise) prior to applying linear multivariate regression techniques. Equivalent concentrations of ampicillin between 0.2mM and 13.5mM in an *E. coli* background could be quantified with good accuracy using this approach.

1. INTRODUCTION

There is a large and continuing interest in the screening of microbial cultures for the production of biologically active metabolites (e.g. [1-18]), which can provide structural templates for synthetic programmes using rational methods of drug design. As well as the increasing use of combinatorial chemical libraries [19-25], methods based on phage display [26-28], synthetic oligonucleotides [29; 30] and DNA shuffling [31-35] can provide further levels of diversity from biological starting points. Recognising the biochemical novelty increasingly being uncovered by genomics [36], modern screens for such metabolites tend to be targeted on the modulation of particular biochemical steps that are thought to be important in the disease process of interest, and can show a high degree of both specificity and sensitivity. This sensitivity means that metabolites showing activity during screening need nowadays be produced only in very small amounts by the organism. In such cases, increasing the titre of the metabolite is vital to provide enough material for further biological evaluation and chemical characterization and, eventually, for commercial production.

The process of titre improvement will normally involve the search for overproducing mutants derived from the original producing organism (see e.g. [6]), but titre-improving

mutants are rare, typically at frequencies of 10^{-4} or less [37], and therefore many thousands of mutants need to be screened in search of an overproducing strain [38]. Previous methods of high-throughput mutant screening have included the assessment of antibiotic activity of the metabolites (e.g. [39]) or use rapid chromatographic methods such as thin layer chromatography (e.g. [40]) or fluorescence and luminescence methods such as the scintillation proximity assay [41-45]. Such methods can typically accommodate 10,000 to 50,000 isolates per month.

The ideal method for culture screening on plates (and indeed for the analysis of fermentor broths generally) would have minimum sample preparation, would analyse samples directly (i.e. be reagentless), would give information about recognizable chemical characters, and would be rapid, automated, noninvasive, quantitative and (at least relatively) inexpensive. These requirements indicate a spectroscopic solution, and we have recently demonstrated that the use of pyrolysis mass spectrometry (PyMS) in combination with a variety of chemometric methods allows rapid screening of cultures for metabolite overproduction [46-48], some 2 min per sample once these have been introduced to the carousel. However, the important conclusion that we would stress is that *whole-cell or whole-broth spectral methods which measure all molecules simultaneously do contain enough spectral information from target molecules of interest to allow their quantification at biotechnologically interesting levels when the entire spectra are used as the inputs to modern chemometric methods based on supervised learning*. The discriminatory power of these chemometric methods is such that in one study [46] we were able to assay quantitatively for the concentrations of ampicillin in mixtures with the Gram-positive *Staphylococcus aureus* when the training set consisted of mixtures of ampicillin and an entirely different biological background, viz. the Gram-negative *Escherichia coli*. This shows, importantly, that chemometric methods of this type, which are designed to effect the quantification of biomolecules in complex biological backgrounds, may indeed be made highly resistant to changes in the background concentrations of metabolites and macromolecles.

As recently reviewed by Magee [49], the chemically-based discrimination of intact microbial cells, referred to as whole-organism fingerprinting, involves the concurrent measurement of large numbers of spectral characters that together reflect the overall cell composition, the commonest spectral approach for this indeed being PyMS. There are, however, four general problems with using PyMS data as the input to supervised learning systems of this type: (i) the method is hardly non-destructive (although this is unimportant for broths, and for plates this could be dealt with by replica plating), (ii) it does not lend itself to *in situ* measurements, (iii) it still suffers somewhat from spectral drift (although recent advances suggest that this problem may be overcome [50; 51]), and (iv) data acquisition still requires nearly 2 minutes per sample. Recently, a number of studies [52-57] have illustrated how even *visible* spectroscopy of petri plates could be used to identify colonies with high levels of electron transport chain components of interest, though this would not of course work directly for most target molecules. Most importantly, however, just as has been widely done with PyMS, Naumann and coworkers in particular (e.g. [58-64]) have shown that FT-IR *absorbance* spectroscopy (in the mid-IR range, defined by IUPAC as 4000-200 cm^{-1} = 2.5-50 μm) provides a powerful tool with sufficient resolving power to distinguish intact microbial cells at the strain level.

In view of the above, we therefore considered that the combination of FT-IR and *supervised* learning methods would, when applied judiciously [65], permit us to extract the chemical concentration of the substance of interest, in a similar manner to that which we developed with PyMS. Sample preparation for absorbance measurements on biological samples of this type is rather tedious, however. Instead, and because FT-IR may be carried out using *reflectance* methods, we considered that one should seek to obtain spectra as a function of spatial location, and by combining the spectroscopy with supervised learning methods obtain images in which metabolite concentrations are encoded as colours or contours, i.e. to construct *a metabolic microscope*. In this regard, it is particularly noteworthy that White's group [66; 67] have shown the ability of diffuse reflectance FT-IR (DRIFT) spectroscopy without any chemometric processing, to effect the discrimination of microbes on surfaces. In a related vein, Yan and colleagues [68] recently showed that FT-IR could be used to analyse solid-state pins as used in combinatorial chemistry, whilst Gremlich and Berets [69] used FT-IR internal reflection spectroscopy for a similar purpose.

We therefore here describe the realisation of our development of diffuse reflectance/absorbance FT-IR spectroscopy as a quantitative tool for the rapid analysis of all samples of biotechnological and other interest, specifically by exploiting the ability of modern, *supervised* learning methods to take multivariate spectral inputs and map them directly to the concentration of one or more target determinands (see above and [70]), using as before [46] mainly mixtures of ampicillin and *E. coli* as a model system.

2. EXPERIMENTAL

2.1 Preparation of the ampicillin mixture with *Escherichia coli*.

E. coli HB101 [71] was used; this is ampicillin-sensitive, indicating that any spectral features observed are not due for instance to ß-lactamase activity. The mixtures were prepared as previously [46]. The strain was grown in 4l liquid medium (glucose (BDH), 10.0g; peptone (LabM), 5.0g; beef extract (LabM), 3.0g; per litre water) for 16h at 37°C in a shaker. After growth the cultures were harvested by centrifugation, washed and resuspended in saline. Ampicillin (desiccated D[-]-α-aminobenzylpenicillin sodium salt, ≧98% (titration), Sigma) was prepared in the bacterial suspensions to give concentration ranges of 0-13.46mM (0 to 5000 μg ml^{-1} in 250 μg ml^{-1} steps) in 40mg ml^{-1} *E. coli* (dry weight) and 0 to 2mM in 200μM steps in 3mg ml^{-1} *E. coli* (dry weight).

2.2 Diffuse reflectance-absorbance FT-IR

20μl aliquots of the above samples were evenly applied onto a flat, sand-blasted aluminium plate (measuring 10cm by 10 cm) and dried at 50°C for 30 min. The plate was mounted onto a motorised stage and the samples analysed using a diffuse reflectance TLC accessory [72-74] connected to a Bruker IFS28 FT-IR spectrometer (Bruker Spectrospin Ltd., Banner Lane, Coventry CV4 9GH, U.K.) equipped with a liquid N_2-cooled MCT (mercury-cadmium-telluride) detector. A schematic of the general optical arrangement of this accessory is shown in Fig 1.

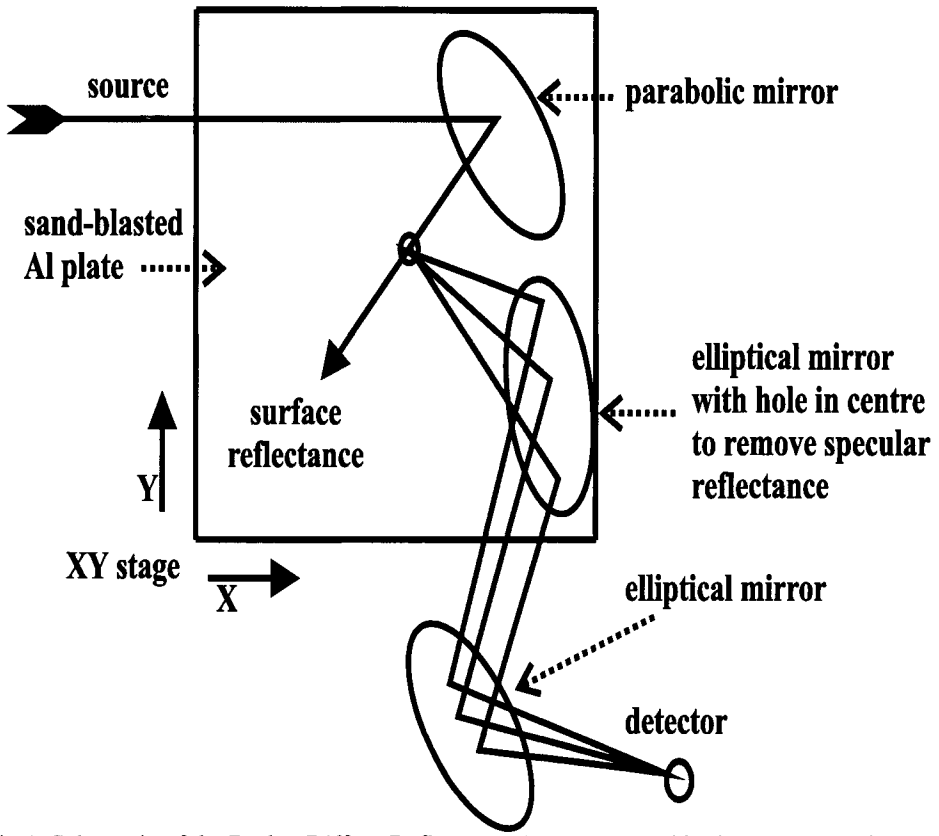

Fig 1. Schematic of the Bruker Diffuse Reflectance Accessory used in the present work

The IBM-compatible PC used to control the IFS28 was programmed (using OPUS version 2.1 software running under IBM OS/2 Warp provided by the manufacturers) to collect spectra over the wavenumber range 4000 cm^{-1} to 600 cm^{-1}. Spectra were acquired at a rate of 20.s^{-1}. The spectral resolution used for the higher concentration experiments was 8 cm^{-1} (whilst the data point spacing in the Fourier transform of the interferogram (after using a zero-filling factor of 2) was 4 cm^{-1}), and for the lower concentration experiments was 4 cm^{-1} (as was the data point spacing in the Fourier transform of the interferogram using a zero-filling factor of 1). To improve the initial signal-to-noise ratio at least 32 spectra were co-added and averaged. Each sample was represented by a spectrum containing 882 points, and spectra were displayed in terms of absorbance as calculated from the reflectance-absorbance spectra using the Opus software and Kubelka-Munk theory. ASCII data were exported from the Opus software used to control the FT-IR instrument and imported into Matlab version 4.2c.1 (The MathWorks, Inc., 24 Prime Park Way, Natick, MA, USA), which runs under Microsoft Windows NT on an IBM-compatible PC.

2.3 PCA and PLS

Matlab was used to perform Principal Components Analysis (PCA) according to the NIPALS algorithm [75], so that exploratory data analysis could be conducted. PCA is a

multivariate statistical technique which can be used to identify *correlations* amongst a set of variables (in this case 882 wavenumbers) and to transform the original set of variables to a new set of *uncorrelated* variables called principal components (PCs). The objective of PCA is to see if the first few PCs account for most (>90%) of the variation in the original data [76]. If they do reduce the number of dimensions required to display the observed relationships, then the PCs can more easily be plotted and 'clusters' in the data visualized [77]; moreover this technique can be used to detect outliers [78]. Matlab routines were also used to perform PLS modelling [79]. PLS is a multivariate technique similar to PCA, but with the components extracted using both x- and y-data and then regressed onto the (known) training results while forming the model. This results in a more parsimonious model in situations where the variance of interest may not be the largest variance in the samples. Data were mean-centred and scaled to unit variance prior to the performance of PCA and PLS, and cross-validation was performed on the test set via the leave-one-out method.

2.4 Artificial neural networks

All artificial neural network (ANN) analyses were carried out with a user-friendly, neural network simulation program, NeuFrame version 1,1,0,0 (Neural Computer Sciences, Lulworth Business Centre, Nutwood Way, Totton, Southampton, Hants), which runs under Microsoft Windows NT on an IBM-compatible PC. In-depth descriptions of the *modus operandi* of this type of ANN analysis [80; 81] in our hands [46-48; 50; 74; 82-92] are given elsewhere.

3. RESULTS AND DISCUSSION

Many studies on the quantification of particular determinands in mixtures, using FTIR, have been based on the contribution of only one or a few spectral features. The carbonyl bond in the β-lactam ring of ampicillin displays a characteristic marker band in the IR spectrum at ~1767 cm^{-1} [93] and, in theory, this property could be used to quantify the concentration of the antibiotic where a good signal-to-noise ratio exists. (The contribution of the carbonyl bond to the spectral pattern was confirmed by incubation of the ampicillin with β-lactamase resulting in β-lactam ring cleavage and a reduction of the absorbance in this region (data not shown).) However, closer analysis of the region between 1850-1700 cm^{-1} indicated that too much baseline variability occurs between individual samples to derive an accurate linear relationship between the absorbance at 1767 cm^{-1} and the known ampicillin concentrations (data not shown). Analysis of the whole spectrum from 4000-600 cm^{-1} indicated that variation in the baseline was evident throughout the whole measured range. Tests on uncoated plates and areas of plates uniformly coated with determinand indicated that this variation was in the background reflectance of the plate *per se*. However, a chemical images of these plates based on integrating the peak at 1767 cm^{-1} is given in Fig 2.

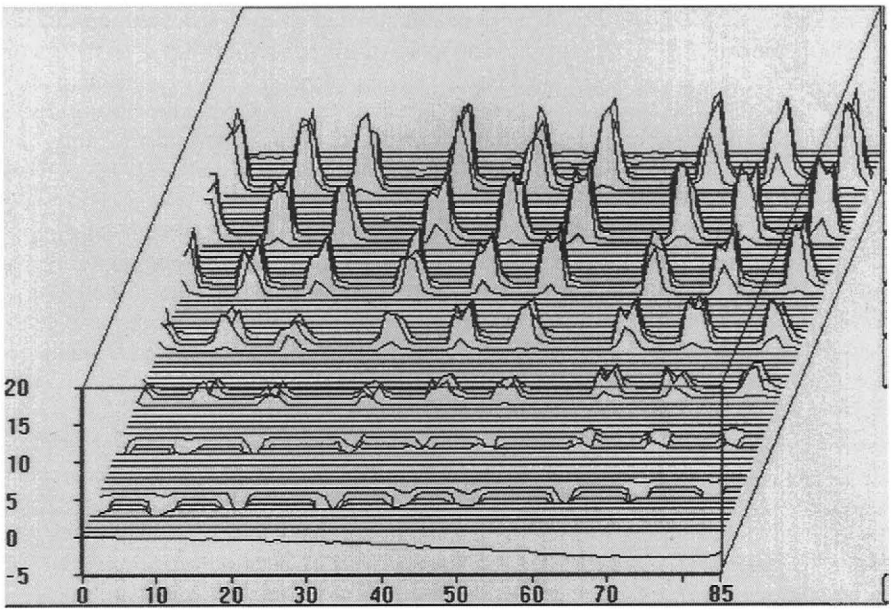

Fig 2. Spatial variation in ampicillin concentration in mixtures with *E. coli* on a diffusely reflecting metal plate. The abscissa is the distance in mm, whilst the ordinate, based on integrating the peak at 1767 cm^{-1}, is arbitrary. Ampicillin concentration increases from lower left to upper right.

Although in this case single spectral features could not be used for accurate quantification of the ampicillin, the implementation of modern chemometric techniques such as PLS and ANNs allowed us to predict the ampicillin concentration using full-spectrum calibration. PLS analysis [79; 94-96] on the full spectral results from the 0-5000μg ml^{-1} ampicillin/*E. coli* samples, where the data were split into a training set (0-5000μg ml^{-1} ampicillin in 500μg ml^{-1} steps) and a test set (250μg ml^{-1}-4750μg ml^{-1} ampicillin in 500μg ml^{-1} steps), resulted in an RMSEP of 6.66% for 7 factors (data not shown). We then used PCA to reduce the number of input nodes to 9 [74]. Such a network trained on a 9-4-1 architecture gave an RMSEP of 3.49% compared to 7.1% for 882-10-1.

Multivariate analysis using PLS and ANNs clearly shows that it is possible to form a model capable of discriminating and quantifying unknown concentrations of ampicillin between 0.25 and 5mg ml^{-1} (0.67-13.46mM) from an *E. coli* cell background. To test the sensitivity of the FTIR/chemometrics approach in determining lower levels of ampicillin we used a concentration range from 0-2mM ampicillin with 3mg ml^{-1} *E. coli* cells in a similar experiment. The RMSEP for the unprocessed spectra was 18.66%; however, this error was improved by the use [61] of the first or second derivatives for input into the PLS (9.4 and 11% for 6 factors).

In addition to the true signal, the raw spectral data also contain noise and baseline shifts which may be derived from instrumental drift (and, in particular here, from variations in

the background reflectivity of the metal plates). Processing the IR spectrum through another level of transformation, from the spectral wavelength domain into the Fourier Domain spectrum (FDS) (or delay domain spectrum), can allow isolation of the signal from the baseline and random (homoscedastic) noise information [97-103]. Whereas the noise is spread throughout the original spectrum, it appears in the high-delay region of the FDS while the signal is concentrated into the low-delay region and the baseline information into the very-low-delay domain. By selecting those variables in the transformed spectrum that correspond most closely to the signal region(s), the majority of the noise and baseline can be removed. Variable selection methods are an extremely powerful adjunct to our hyperspectral approaches [65; 104; 105], and the same RMSEP may often be acquired from a very small fraction of the variables (e.g. wavelengths) available [106]. Not only can the effective removal of these variables in the PLS model improve the PLS prediction but such parimonious models are widely considered to be more robust and to generalise better [107].

PLS performed on the Fourier Domain spectra (FDS) (or delay domain spectra) produced an RMSEP of 4.12% with 7 regression factors. PLS similarly performed on the FDS from the 0-5000μg ml^{-1} ampicillin data set produced an RMSEP of 4.28% (8 factors; data not shown).

Finally, we have used cognate methods to effect the sensitive discrimination of various enterococci and streptococci [74] and to provide accurate, quantitative estimation of aristeromycin and neplanocin A (see e.g. [108]) in *S. citricolor* fermentations; a full description of the latter is given in the paper by Winson *et al.* elsewhere in this volume. However, an important point to be made is that estimations in whole fermentor broths and the like differ from those in simple mixtures in that the chemometric methods can exploit differences in the *organism* or the *medium* which *correlate* with metabolite overproduction, rather than relying solely on spectral features due to the target molecules *themselves*. This can serve to give an extremely useful 'amplification' to the method when the target concentrations are partiocularly low.

4. CONCLUSIONS

Driven in part by the activities of the "remote sensing" community [109], there is much interest in the rapid acquisition of diffuse reflectance spectral data from various spatial locations on the earth, detecting hundreds of wavelengths simultaneously (most commonly in the visible and near infrared), and coupled increasingly to advanced data reduction and visualization algorithms, an approach that is nowadays often referred to [110-118] as hyperspectral imaging. Such remote-sensing analyses occasionally use the mid-IR part of the spectrum [119-122], but one of the problems with this approach to *remote* sensing is the strong and variable absorbance of radiation by the atmosphere itself [118; 123; 124], a problem from which we do not really suffer.

Diffuse reflectance FTIR in combination with a multivariate calibration chemometric approach to data analysis could be used to effect the rapid quantification of a pharmaceutical product (ampicillin) in a (variable) biological background (*E. coli* cells), a situation representative of metabolite over-production in a screening or titre improvement programme [46]. Spectral variation contributed by shifting baseline due to instrumental interference and differences in the biological background between samples (which would have prevented

accurate univariate calibration) could largely be eliminated by PCA (as seen previously [74]) or by transformation to the Fourier domain prior to forming a model with ANNs or PLS. Although diffuse reflectance methods are well known to suffer difficulties in traditional quantitative work (the concentration region for which Kubelka-Munk theory [125-128] holds, for instance, is normally quite small (see e.g. [129]), it is clear that the combination of modern chemometric methods with the diffuse reflectance-absorbance approach overcomes these most satisfactorily. Thus we have here shown for the first time that the hyperspectral approach using diffuse reflectance-absorbance spectroscopy, when coupled to modern supervised learning methods, provides a novel, rapid, general and powerful approach to the problem of screening for metabolite overproduction in biological and biotechnological systems.

ACKNOWLEDGEMENTS

DBK, MKW, AMW, BKA, AUJ and JJR thank the Chemicals and Pharmaceuticals Directorate of the UK BBSRC, GlaxoWellcome and Bruker Spectrospin Ltd for financial support, whilst RG and EMT thank the Wellcome Trust (grant number 042615/Z/94/Z) for the same reason. We thank Drs Paul Turner (Bruker), Tony Buss, Mike Dawson, Brian Rudd and Martin Todd (all GlaxoWellcome) for a number of useful discussions.

REFERENCES

1. Bu'Lock, J. D., Nisbet, L. J. & Winstanley, D. J. (1982)*Bioactive microbial products: search and discovery*. Academic, New York.
2. Elander, R. P. (1987) Microbial screening, selection and strain improvement. In *Basic Biotechnology* (ed. J. Bu'lock and B. Kristiansen), pp. 217-251. Academic Press, London.
3. Vandamme, E. J. (1984) *Biotechnology in industrial antibiotics*. Marcel Dekker, New York.
4. Verrall, M. S. (1985) Discovery and isolation of microbial products. Ellis Horwood, Chichester.
5. Ōmura, S. (1986) Philosophy of new drug discovery. *Microbiol. Rev.* **50,** 259-279.
6. Crueger, W. & Crueger, A. (1989) *Biotechnology: a textbook of industrial microbiology*. Sinauer Associates, Sunderland, MA.
7. Porter, N. & Fox, F. (1993) Diversity of microbial products and applications. *Pesticide Sci.* **39,** 161-168.
8. Nisbet, L. J. & Porter, N. (1989) The impact of pharmacology and molecular biology on the exploitation of natural products. *Symp. Soc. Gen. Microbiol.* **44,** 309-342.
9. Crittenden, P. D. & Porter, N. (1991) Lichen-forming fungi - potential sources of novel metabolites. *Trends Biotechnol.* **9,** 409-414.
10. Franco, C. M. M. & Countinho, L. E. L. (1991) Detection of novel secondary metabolites. *CRC Crit. Rev. Biotechnol.* **11,** 193-276.
11. Lancini, G. & Lorenzetti, R. (1993) *Biotechnology of antibiotics and other bioactive microbial metabolites*. Plenum Press, New York.
12. Bevan, P., Ryder, H. & Shaw, A. (1995) Identifying small-molecule lead compounds - the screening approach to drug discovery. *Trends Biotechnol.* **13,** 115-121.

13. Babiak, J., Lucotch, B., Russo, A., Heydt, L., Williams, S. & McCaully, R. (1995) The Trials and Tribulations of a Robotic Screening Core. *J. Automatic Chem.* **17,** 55-58.
14. Lin, B. B. (1995) High-throughput screening for new drug discovery. *J. Food Drug Anal.* **3,** 233-241.
15. Ashton, M. J., Jaye, M. & Mason, J. S. (1996) New Perspectives in Lead Generation .1. Discovery of Biological Targets. *Drug Discovery Today* **1,** 11-15.
16. Ashton, M. J., Jaye, M. C. & Mason, J. S. (1996) New Perspectives in Lead Generation .2. Evaluating Molecular Diversity. *Drug Discovery Today* **1,** 71-78.
17. Broach, J. R. & Thorner, J. (1996) High-throughput screening for drug discovery. *Nature* **384 Supplement,** 14-16.
18. Palmer, M. A. J. (1996) High-throughput screening on a low budget. *Nature Biotechnology* **14,** 513-515.
19. Terrett, N. K., Gardner, M., Gordon, D. W., Kobylecki, R. J. & Steele, J. (1995) Combinatorial synthesis - the design of compound libraries and their application to drug discovery. *Tetrahedron* **51,** 8135-8173.
20. Gallop, M. A., Barrett, R. W., Dower, W. J., Fodor, S. P. A. & Gordon, E. M. (1994) Applications of combinatorial technologies to drug discovery .1. Background and peptide combinatorial libraries. *J. Med. Chem.* **37,** 1233-1251.
21. De Witt, S. H., Schroeder, M. C., Stankovic, C. J., Strode, J. E. & Czarnik, A. W. (1994) Diversomer™ technology - solid-phase synthesis, automation, and integration for the generation of chemical diversity. *Drug Development Research* **33,** 116-124.
22. Gordon, E. M., Barrett, R. W., Dower, W. J., Fodor, S. P. A. & Gallop, M. A. (1994) Applications of combinatorial technologies to drug discovery. 2. Combinatorial organic synthesis, library screening strategies, and future directions. *J. Med. Chem.* **37,** 1385-1401.
23. Blondelle, S. E., Perezpaya, E., Dooley, C. T., Pinilla, C. & Houghten, R. A. (1995) Soluble Combinatorial Libraries of Organic, Peptidomimetic and Peptide Diversities. *Trac-Trends in Analytical Chemistry* **14,** 83-92.
24. Ecker, D. J. & Crooke, S. T. (1995) Combinatorial drug discovery: which methods will produce the greatest value? *Bio/Technol.* **13,** 351-360.
25. Janda, K. D. (1994) Tagged versus untagged libraries - methods for the generation and screening of combinatorial chemical libraries. *Proc. Natl. Acad. Sci.* **23,** 10779-10785.
26. Clackson, T., Hoogenboom, H. R., Griffiths, A. D. & Winter, G. (1991) Making antibody fragments using phage display libraries. *Nature* **352,** 624-628.
27. Lowman, H. B., Bass, S. H., Simpson, N. & Wells, J. A. (1991) Selecting high-affinity binding-proteins by monovalent phage display. *Biochemistry* **30,** 10832-10838.
28. Garrard, L. J., Yang, M., O'Connell, M. P., Kelley, R. F. & Henner, D. J. (1991) Fab assembly and enrichment in a monovalent phage display system. *Bio/Technology* **9,** 1373-1377.
29. Needels, M. C., Jones, D. G., Tate, E. H., Heinkel, G. L., Kocher Sperger, L. M. & Dower, W. J. (1993) Generation and screening of an oligonucleotide-encoded synthetic peptide library. *Proc. Natl. Acad. Sci.* **90,** 10700-10704.
30. Lipshutz, R. J., Morris, D., Chee, M., Hubbell, E., Kozal, M. J., Shah, N., Shen, N., Yang, R. & Fodor, S. P. A. (1995) Using oligonucleotide probe arrays to access genetic diversity. *Biotechniques* **19,** 442-447.
31. Stemmer, W. P. C. (1994) Rapid evolution of a protein in *vivo* by DNA shuffling. *Nature* **370,** 389-391.

32. Stemmer, W. P. C. (1994) DNA shuffling by random fragmentation and reassembly: in *vitro* recombination for molecular evolution. *Proc. Natl. Acad. Sci. USA* **91,** 10747-10751.
33. Stemmer, W. P. C., Crameri, A., Ha, K. D., Brennan, T. M. & Heyneker, H. L. (1995) Single-step assembly of a gene and entire plasmid from large numbers of oligodeoxyribonucleotides. *Gene* **164,** 49-53.
34. Crameri, A., Cwirla, S. & Stemmer, W. P. C. (1996) Construction and evolution of antibody phage libraries by DNA shuffling. *Nature Medicine* **2,** 100-102.
35. Crameri, A., Whitehorn, E. A., Tate, E. & Stemmer, W. P. C. (1996) Improved green fluorescent protein by molecular evolution using DNA shuffling. *Nature Biotechnology* **14,** 315-319.
36. Oliver, S. G. (1996) From DNA sequence to biological function. *Nature* **379,** 597-600.
37. Rowlands, R. T. (1984) Industrial strain improvement: mutagenesis and random screening procedures. *Enz. Micr. Technol.* **6,** 3-10.
38. An, G., Bielich, J., Auerbach, R. & Johnson, E. A. (1991) Isolation and characterization of carotenoid hyperproducing mutants of yeast by flow cytometry and cell sorting. *Bio/Technology* **9,** 69-73.
39. Ball, C. & McGonagle, M. P. (1987) Development and evaluation of a potency index screen for detecting mutants of *Penicillium chrysogenum* having increased penicillin yield. *J. Appl. Bacteriol.* **45,** 67-74.
40. Spagnoli, R. & Cappelletti, L. (1981) New screening method for rapid identification of blocked mutants: application to erythromycin- and oleandomycin-producing strains. *Agric. Biol. Chem.* **45,** 761-763.
41. Bosworth, N. & Towers, P. (1989) Scintillation Proximity Assay. *Nature* **341,** 167-168.
42. Nelson, N. (1987) A novel method for the detection of receptors and membrane proteins by scintillation proximity radioassay. *Analytical Biochemistry* **165,** 287-293.
43. Udenfriend, S., Gerber, L. & Nelson, N. (1987) Scintillation proximity assay - a sensitive and continuous isotopic method for monitoring ligand receptor and antigen-antibody interactions. *Analytical Biochemistry* **161,** 494-500.
44. Udenfriend, S., Gerber, L. D., Brink, L. & Spector, S. (1985) Scintillation proximity radioimmunoassay utilizing I-125-labeled ligands. *Proc. Natl. Acad. Sci.* **82,** 8672-8676.
45. Major, J. S. (1995) Challenges of High-Throughput Screening Against Cell-Surface Receptors. *Journal of Receptor and Signal Transduction Research* **15,** 595-607.
46. Goodacre, R., Trew, S., Wrigley-Jones, C., Neal, M. J., Maddock, J., Ottley, T. W., Porter, N. & Kell, D. B. (1994) Rapid screening for metabolite overproduction in fermentor broths, using pyrolysis mass spectrometry with multivariate calibration and artificial neural networks. *Biotechnol. Bioeng.* **44,** 1205-1216.
47. Goodacre, R., Trew, S., Wrigley-Jones, C., Saunders, G., Neal, M. J., Porter, N. & Kell, D. B. (1995) Rapid and quantitative analysis of metabolites in fermentor broths using pyrolysis mass spectrometry with supervised learning: application to the screening of *Penicillium chryosgenum* fermentations for the overproduction of Penicillins. *Anal. Chim. Acta* **313,** 25-43.
48. Goodacre, R. & Kell, D. B. (1996) Pyrolysis mass spectrometry and its applications in biotechnology. *Curr. Opinion Biotechnol.* **7,** 20-28.
49. Magee, J. T. (1993) Whole-organism fingerprinting. In *Handbook of New Bacterial Systematics* (ed. M. Goodfellow and A. G. O'Donnell), pp. 383-427. Academic Press, London.

50. Goodacre, R. & Kell, D. B. (1996) Correction of mass spectral drift using artificial neural networks. *Anal. Chem.* **68,** 271-280.
51. Goodacre, R., Timmins, É. M., Jones, A., Kell, D. B., Maddock, J., Heginbothom, M. L. & Magee, J. T. (1996) On mass spectrometer instrument standardization and interlaboratory calibration transfer using neural networks. *Anal. Chim. Acta*, submitted.
52. Yang, M. M. & Youvan, D. C. (1988) Applications of imaging spectroscopy in molecular biology. 1. Screening photosynthetic bacteria. *Bio/Technology* **6,** 939-942.
53. Arkin, A. P., Goldman, E. R., Robles, S. J., Goddard, C. A., Coleman, W. J., Yang, M. M. & Youvan, D. C. (1990) Applications of imaging spectroscopy in molecular biology .2. Colony screening based on absorption spectra. *Bio/Technology* **8,** 746-749.
54. Goldman, E. R. & Youvan, D. C. (1992) An algorithmically optimized combinatorial library screened by digital imaging spectroscopy. *Bio/Technology* **10,** 1557-1561.
55. Arkin, A. P. & Youvan, D. C. (1993) Digital imaging spectroscopy. In *The photosynthetic reaction center, Vol 1.* (ed. J. Deisenhofer and J. R. Norris), pp. 133-155. Academic Press, New York.
56. Youvan, D. (1994) Imaging sequence space. *Nature* **369,** 79-80.
57. Brown, S., Colson, A.-M., Meunier, B. & Rich, P. R. (1993) Rapid Screening of Cytochromes of respiratory Mutants of *Saccharomyces cerevisiae. Eur. J. Biochem.* **213,** 137-145.
58. Helm, D., Labischinski, H. & Naumann, D. (1991) Elaboration of a procedure for identification of bacteria using Fourier-transform IR spectral libraries: a stepwise correlation approach. *Microbiological methods* **14,** 127-142.
59. Helm, D., Labischinski, H., Schallehn, G. & Naumann, D. (1991) Classification and identification of bacteria by Fourier transform infrared spectroscopy. *J. Gen. Microbiol.* **137,** 69-79.
60. Naumann, D. (1985) The ultra rapid differentiation and identification of pathogenic bacteria using FT-IR techniques. *Fourier and computerized infrared spectroscopy* **553,** 268-269.
61. Naumann, D., Fijala, V. & Labischinski, H. (1988) The differentiation and identification of pathogenic bacteria using FT-IR and multivariate statistical analysis. *Mikrochim Acta* **1,** 373-377.
62. Naumann, D., Fijjala, V., Lavischinski, H. & Giesbrecht, P. (1988) The rapid differentiation and identification of pathogenic bacteria using fouriter transform infrared spectroscopic and multivariate statistical analysis. *Molecular structure* **174,** 165-170.
63. Naumann, D., Helm, D. & Labischinski, H. (1991) Microbiological characterizations by FT-IR spectroscopy. *Nature* **351,** 81-82.
64. Naumann, D., Keller, S., Helm, D., Schultz, C. & Schrader, B. (1995) FT-IR spectroscopy and FT-Raman spectroscopy are powerful analytical tools for the noninvasive characterization of intact microbial cells. *Journal of Molecular Structure* **347,** 399-405.
65. Kell, D. B. & Sonnleitner, B. (1995) GMP - Good Modelling Practice: an essential component of good manufacturing practice. *Trends Biotechnol.* **13,** 481-492.
66. Nichols, P. D., Henson, J. M., Guckert, J. B., Nivens, D. E. & White, D. C. (1984) Fourier transform infrared spectroscopic methods for microbial ecology: analysis of bacteria bacteria-polymer mixtures and biofilms. *J. Microbiological methods* **4,** 79-94.
67. Nivens, D. E., Chambers, J. Q., Anderson, T. R., Tunlid, A., Smit, J. & White, D. C. (1993) Monitoring microbial adhesion and biofilm formation by attenuated total

reflection/fourier transform infrared spectroscopy. *J. Microbiological methods* **17,** 199-213.

68. Yan, B., Kumaravel, G., Anjaria, H., Wu, A., Petter, R. C., Jewell Jr., C. F. & Wareing, J. R. (1995) Infrared spectrum of a single resin bead for real-time monitoring of solid-phase reactions. *J. Org. Chem.* **60,** 5736-5738.
69. Gremlich, H.-U. & Berets, S. L. (1996) Use of FT-IR internal reflection spectroscopy in combinatorial chemistry. *Appl. Spectrosc.* **50,** 532-536.
70. Zupan, J. & Gasteiger, J. (1993) *Neural Networks for Chemists.* Verlag Chemie, Weinheim.
71. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular cloning: a laboratory manual.* Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
72. Bouffard, S. P., Katon, J. E., Sommer, A. J. & Danielson, N. D. (1994) Development of microchannel thin layer chromatography with infrared microspectroscopic detection. *Analytical chemistry* **66,** 1937-1940.
73. Glauninger, G., Kovar, K. A. & Hoffmann, V. (1990) Possibilities and limits of an online coupling of thin-layer chromatography and FTIR spectroscopy. *Fresenius J. Anal. Chem* **338,** 710-716.
74. Goodacre, R., Timmins, É. M., Rooney, P. J., Rowland, J. J. & Kell, D. B. (1996) Rapid identification of *Streptococcus* species using diffuse reflectance-absorbance Fourier transform infrared spectroscopy and artificial neural networks. *FEMS Microbiol. Lett.* **140,** 233-239.
75. Wold, H. (1966) Estimation of principal components and related models by iterative least squares. In *Multivariate Analysis* (ed. K. R. Krishnaiah), pp. 391-420. Academic Press, New York.
76. Hotelling, H. (1933) Analysis of a Complex of Statistical Variables into Principal Components. *Journal of Educational Psychology* **24,** 417-441 and 498-520.
77. Everitt, B. S. (1993) *Cluster Analysis.* Edward Arnold, London.
78. Jolliffe, I. T. (1986) *Principal Component Analysis.* Springer-Verlag, New York.
79. Martens, H. & Næs, T. (1989) *Multivariate calibration.* John Wiley, Chichester.
80. Rumelhart, D. E., McClelland, J. L. & The PDP Research Group (1986) Parallel Distributed Processing. Experiments in the Microstructure of Cognition, Vols I & II. M.I.T. Press, Cambridge, MA.
81. Werbos, P. J. (1994) *The roots of back-propagation: from ordered derivatives to neural networks and political forecasting.* John Wiley, Chichester.
82. Goodacre, R. (1992) The effects of genotypic and phenotypic changes on bacterial identification using pyrolysis mass spectrometry. PhD thesis, University of Bristol.
83. Goodacre, R. & Kell, D. B. (1993) Rapid and quantitative analysis of bioprocesses using pyrolysis mass spectrometry and neural networks. Application to indole production. *Analytica Chimica Acta* **279,** 17-26.
84. Goodacre, R., Kell, D. B. & Bianchi, G. (1993) Rapid assessment of the adulteration of virgin olive oils by other seed oils using pyrolysis mass spectrometry and artificial neural networks. *Journal of the Science of Food and Agriculture* **63,** 297-307.
85. Goodacre, R., Edmonds, A. N. & Kell, D. B. (1993) Quantitative analysis of the pyrolysis mass spectra of complex mixtures using artificial neural networks. Application to amino acids in glycogen. *Journal of Analytical and Applied Pyrolysis* **26,** 93-114.

86. Goodacre, R., Neal, M. J. & Kell, D. B. (1994) Rapid and quantitative analysis of the pyrolysis mass spectra of complex binary and tertiary mixtures using multivariate calibration and artificial neural networks. *Analytical Chemistry* **66,** 1070-1085.
87. Goodacre, R., Karim, A., Kaderbhai, M. A. & Kell, D. B. (1994) Rapid and quantitative analysis of recombinant protein expression using pyrolysis mass spectrometry and artificial neural networks - application to mammalian cytochrome b_5 in *Escherichia coli*. *Journal of Biotechnology* **34,** 185-193.
88. Goodacre, R. (1994) Characterization and quantification of microbial systems using pyrolysis mass spectrometry: introducing neural networks to analytical pyrolysis. *Microbiology Europe* **2,** 16-22.
89. Goodacre, R., Neal, M. J., Kell, D. B., Greenham, L. W., Noble, W. C. & Harvey, R. G. (1994) Rapid identification using pyrolysis mass spectrometry and artificial neural networks of *Propionibacterium acnes* isolated from dogs. *J. Appl. Bacteriol.* **76,** 124-134.
90. Goodacre, R. (1994) Characterisation and quantification of microbial systems using pyrolysis mass spectrometry: Introducing neural networks to analytical pyrolysis. *Microbiology Europe* **2,** 16-22.
91. Neal, M. J., Goodacre, R. & Kell, D. B. (1994) On the analysis of pyrolysis mass spectra using artificial neural networks. Individual input scaling leads to rapid learning. In *Proceedings of the World Congress on Neural Networks*, vol. I, PP. 318-323. International Neural Network Society, San Diego.
92. Goodacre, R., Howell, S. A., Noble, W. C. & Neal, M. J. (1996) Sub-species discrimination using pyrolysis mass spectrometry and self-organising neural networks of *Propionibacterium acnes* isolated from normal human skin. *Zentralblatt für Bakteriologie* **284,** 501-515.
93. Wong, J. S., Rein, A. J., Wilks, D. & Wilks, P. (1984) Infrared spectroscopy of aqueous antibiotics solutions. *Applied spectroscopy* **38,** 32-35.
94. Bookstein, F. L. (1980) Data analysis by partial least squares. In *Evaluation of econometric models* (ed. J. Kmenta and J. B. Ramsey), pp. 75-90. Academic Press.
95. Haaland, D. M. & Thomas, E. V. (1988) Partial Least Squares Methods for Spectral Analyses. 1. Relation to Other Quantitative Calibration Methods and the Extraction of Qualitative Information. *Anal. Chem.* **60,** 1193-1202.
96. Höskuldsson, A. (1988) PLS regression methods. *Journal of chemometrics* **2,** 211-228.
97. Arnold, M. A. & Small, G. W. (1990) Determination of physiological levels of glucose in an aqueous matrix with digitally filtered Fourier-transform near-infrared spectra. *Anal. Chem.* **62,** 1457-1464.
98. Small, G. W., Harms, A. C., Kroutil, R. T., Ditillo, J. T. & Loerop, W. R. (1990) Design of optimized finite impulse-response digital-filters for use with passive Fourier-transform infrared interferograms. *Anal. Chem.* **62,** 1768-1777.
99. Small, G. W. & Barber, A. S. (1992) Application of digital filtering and pattern recognition techniques to interferogram based Fourier transform infrared qualitative analysis. Investigation of special interferences. *Chemometrics and intelligent laboratory systems* **15,** 203-217.
100. Pan, S. T., Chung, H., Arnold, M. A. & Small, G. W. (1996) Near-Infrared spectroscopic measurement of physiological glucose levels in variable matrices of protein and triglycerides. *Analytical Chemistry* **68,** 1124-1135.
101. Mattu, M. J. & Small, G. W. (1995) Quantitative analysis of bandpass-filtered Fourier-transform infrared interferograms. *Analytical Chemistry* **67,** 2269-2278.

102. Shaffer, R. E., Small, G. W., Combs, R. J., Knapp, R. B. & Kroutil, R. T. (1995) Experimental-Design Protocol For the Pattern-Recognition Analysis of Bandpass Filtered Fourier-Transform Infrared Interferograms. *Chemometrics and Intelligent Laboratory Systems* **29,** 89-108.
103. Woodward, A. M., Alsberg, B. K. & Kell, D. B. (1997) The effect of heteroscedastic noise on the modelling of frequency domain data. *Appl. Spectroscopy,* submitted
104. Goodacre, R., Neal, M. J. & Kell, D. B. (1996) Quantitative analysis of multivariate data using artificial neural networks: a tutorial review and applications to the deconvolution of pyrolysis mass spectra *Zentralblatt. für Bakteriologie* **284,** 516-539.
105. Shaw, A. D., di Camillo, A., Vlahov, G., Kell, D. B., Rowland, J. & Bianchi, G. (1997) Discrimination of variety and region of origin of extra virgin olive oils using ^{13}C NMR and variable reduction. *Anal. Chim. Acta* , in the press.
106. Broadhurst, D., Goodacre, R., Jones, A., Rowland, J. J. & Kell, D. B. (1997) Genetic algorithms as a method for variable selection in PLS regression, with applications to pyrolysis mass spectrometry. *Anal. Chim. Acta* , in the press.
107. Seasholtz, M. B. & Kowalski, B. (1993) The parsimony principle applied to multivariate calibration. *Anal. Chim. Acta.* **277,** 165-177.
108. Jenkins, G. N. & Turner, N. J. (1995) The biosynthesis of carbocyclic nucleosides. *Chemical Society Reviews* **24,** 169-176.
109. Curran, P. J. (1994) Imaging Spectrometry. *Progress in Physical Geography* **18,** 247-266.
110. Chiou, W. C. (1985) NASA image-based geological expert system - development project for hyperspectral image analysis. *Applied Optics* **24,** 2085-2091.
111. Goetz, A. F. H., Vane, G., Solomon, J. & Rock, B. N. (1985) Imaging spectrometry for earth remote sensing. *Science* **228,** 1147-1153.
112. Birk, R. J. & McCord, T. B. (1994) Airborne hyperspectral sensor systems. *IEEE Aerospace and Electronic Systems Magazine* **9,** 26-33.
113. Filiberti, D. P., Marsh, S. E. & Schowengerdt, R. A. (1994) Synthesis of imagery with high spatial and spectral resolution from multiple image sources. *Optical Engineering* **33,** 2520-2528.
114. Abousleman, G. P., Gifford, E. & Hunt, B. R. (1994) Enhancement and compression techniques for hyperspectral data. *Optical Engineering* **33,** 2562-2571.
115. Lee, Z. P., Carder, K. L., Hawes, S. K., Steward, R. G., Peacock, T. G. & Davis, C. O. (1994) Model for the interpretation of hyperspectral remote sensing reflectance. *Applied Optics* **33,** 5721-5732.
116. Rogers, S. K., Colombi, J. M., Martin, C. E., Gainey, J. C., Fielding, K. H., Burns, T. J., Ruck, D. W., Kabrisky, M. & Oxley, M. (1995) Neural networks for automatic target recognition. *Neural Networks* **8,** 1153-1184.
117. Wilson, T. A., Rogers, S. K. & Myers, L. R. (1995) Perceptual-based hyperspectral image fusion using multiresolution analysis. *Optical Engineering* **34,** 3154-3164.
118. Gaddis, L. R., Soderblom, L. A., Kieffer, H. H., Becker, K. J., Torson, J. & Mullins, K. (1996) Decomposition of AVIRIS spectra - extraction of surface reflectance, atmospheric, and instrumental components. *IEEE Transactions on Geoscience and Remote Sensing* **34,** 163-178.
119. Benediktsson, J. A., Sveinsson, J. R. & Arnason, K. (1995) Classification and feature extraction of AVIRIS data. *IEEE Transactions on Geoscience and Remote Sensing* **33,** 1194-1205.

120. Preston, E., Bergman, T., Gorenflo, R., Hermann, D., Kopala, E., Kuzma, T., Lazofson, L. & Orkis, R. (1994) Development of a field-portable imaging system for scene classification using multispectral data fusion algorithms. *IEEE Aerospace and Electronic Systems Magazine* **9,** 13-19.
121. Wienke, D., Vandenbroek, W., Melssen, W., Buydens, L., Feldhoff, R., Kantimm, T., Huthfehre, T., Quick, L., Winter, F. & Cammann, K. (1995) Comparison of an adaptive resonance theory-based neural network (ART- 2A) against other classifiers for rapid sorting of post consumer plastics by remote near-infrared spectroscopic sensing using an InGaAs diode array. *Analytica Chimica Acta* **317,** 1-16.
122. Persky, M. J. (1995) A review of spaceborne infrared Fourier-Transform spectrometers for remote sensing. *Review of Scientific Instruments* **66,** 4763-4797.
123. Vane, G., Duval, J. E. & Wellman, J. B. (1993) Imaging spectroscopy of the earth and other solar system bodies. In *Remote geochemical analysis: elemental and mineralogical composition* (ed. C. M. Pieters and P. A. J. Englert), pp. 121-143. Cambridge University Press, Cambridge.
124. Dwyer, J. L., Kruse, F. A. & Lefkoff, A. B. (1995) Effects of empirical versus model-based reflectance calibration on automated analysis of imaging spectrometer data - a case study from the Drum mountains, Utah. *Photogrammetric Engineering and Remote Sensing* **61,** 1247-1254.
125. Fuller, M. P. & Griffiths, P. R. (1978) Diffuse reflectance measurements by infrared Fourier transform spectroscopy. *Anal. Chem.* **50,** 1906-1910.
126. Griffiths, P. R. & de Haseth, J. A. (1986) *Fourier transform infrared spectrometry.* John Wiley, New York.
127. Mitchell, M. B. (1993) Fundamentals and applications of diffuse reflectance infrared Fourier transform (DRIFT) spectroscopy. *ACS Advances in Chemistry Series* **236,** 351-375.
128. Schrader, B. (1995) *Infrared and Raman spectroscopy: methods and applications.* Verlag Chemie, Weinheim.
129. Hapke, B. (1993) *Theory of reflectance and emittance spectroscopy.* Cambridge University Press., Cambridge.