Reprinted from **Clinical Flow Cytometry**
Volume 677 of the *Annals of the New York Academy of Sciences*
March 20, 1993

White Cell and Thrombocyte Disorders

Standardized, Self-learning Flow Cytometric List Mode Data Classification with the CLASSIF1 Program System

G. VALET,[a] M. VALET,[b] D. TSCHÖPE,[c] H. GABRIEL,[d]
G. ROTHE,[a] W. KELLERMANN,[e] AND H. KAHLE[a]

[a]*Max-Planck-Institut für Biochemie*
D-8033 Martinsried, Germany

[b]*Max-Planck-Institut für Biochemie*
D-8000 München, Germany

[c]*Diabetes Forschungsinstitut der Heinrich-Heine-Universität*
D-4000 Düsseldorf, Germany

[d]*Institut für Sport- und Leistungsmedizin der Universität*
D-6600 Saarbrücken, Germany

[e]*Institut für Anästhesiologie*
Klinikum Großhadern der Universität
D-8000 München, Germany

INTRODUCTION

White blood cells are mediators of systemic cellular functions like immune defense for lymphocytes and monocytes, phagocytosis and digestion of microorganisms for granulocytes, and primary hemostasis for trombonists. The immunological and biochemical features of these cells can be flow-cytometrically determined by (i) immunophenotyping for lymphocytes and monocytes, (ii) the measurement of functional reactions like phagocytosis, killing of bacteria, intracellular calcium levels, and respiratory burst or protease activities, or (iii) the expression of activation antigens for granulocytes, monocytes, and trombonists.

Although one or several multiparametric flow cytometric measurements on any one of these cell types provide a high degree of information, the computer-assisted extraction of the characteristic values of two-dimensional or multidimensional cell clusters such as means, coefficients of variation, modes, or medians as the most widespread method of analysis does not sufficiently extract this information.

Cluster analysis,[1–6] fitting of multidimensional Gaussians,[7,8] principal component/biplot analysis,[9,10] use of neural networks[11–13] or expert systems,[14–17] and multiwindow content evaluation[18,19] have been proposed for automated result extraction. However, progress in view of a generally and routinely applicable automated flow cytometric list mode classification program system has remained quite limited.

The newly developed multiwindow evaluation program system CLASSIF1 presents a practical way for a standardized, generally applicable, self-learning and simultaneous classification of multiple multiparameter measurements. The classifica-

tions are based on observed data. No assumptions with regard to parameter distribution functions are made and unsupervised self-learning is performed.

The following examples of antigenic and functional cell characterization demonstrate the utility of this general list mode data evaluation concept for (i) the recognition of abnormal trombonists in patients with risk for myocardial infarction and manifest diabetes,[20–24] (ii) the evaluation of lymphocyte phenotypes from normal and endurance-trained competition cyclists suffering from an overtraining syndrome,[25–30] and (iii) the recognition of infection and sepsis in intensive-care-unit (ICU) patients.[31]

MATERIALS AND METHODS

Thrombocyte Immunophenotyping (Düsseldorf III Activation Marker Test)

Patients with diabetes type I insulin-dependent diabetes mellitus (IDDM) or type II noninsulin-dependent diabetes mellitus (NIDDM) and patients with clinically manifest, angiographically proven high-grade coronary heart disease (CHD) as high-risk patients for myocardial infarction were investigated,[22,23] whereas healthy volunteers without any drug medication and with negative smoking status served as normal controls.

Fresh venous blood (4.5 mL) was anticoagulated either with 0.5 mL citrate solution containing 0.13 M Na-citrate (Merck 9361, Darmstadt, Germany), 0.05 M $Na/K/HPO_4^{3-}$, 0.034 M NaCl, 55.5 μM Aspisol (Bayer, Germany), and 30 μM prostaglandin E1 (Sigma, St. Louis, Missouri) or with 0.134 M EDTA solution containing 7 mg/mL hydroxychloroquin-sulfate and 20 IU/mL heparin. This was subsequently mixed in a proportion of 1:1 (v/v) with a 1.0% formaldehyde solution in PBS of pH 7.2 and incubated for 15 min at room temperature (RT), followed by centrifugation at 100*g* for 10 min at RT. The supernatant was removed as platelet-rich plasma (PRP) and centrifuged at 700*g* for 5 min at RT, followed by resuspension of the pellet in 1 mL of a 0.13 M Na-tricitrate (Merck 6448) solution. This PRP suspension was diluted to a thrombocyte concentration of 5×10^4/mL and 200 μL of thrombocyte suspension was incubated at RT for 1 h at 90% antibody saturation with 50 μL of plateau-titered monoclonal antithrombospondin P10 (Dianova, Hamburg, Germany) and CD62 [2.17] or CD63 [2.28] antibodies [kindly provided by Nieuwenhuis *et al.* (Utrecht, the Netherlands) or obtained from Dianova]. The unspecific isotype-matched IgG was from Coulter (Hialeah, Florida). The assays were developed for 30 min at RT in the dark by addition of 100 μL of titered anti-mouse $F(ab)_2$-FITC antibodies (Sigma), followed by addition of 1 mL of washing solution and centrifugation at 700*g* for 5 min at RT. The supernatant was diluted 1:1 (v/v) with sheath fluid [Becton Dickinson (BD), Heidelberg, Germany] and was measured in a FACS 440 cell sorter (BD).

Four-decade logarithmically amplified forward (FSC) and sideward (SSC) scatter signals and FITC fluorescence signals of individual thrombocytes were collected and stored as FCS1.0 list mode files. Fluorescence signal intensity and discrimination were calibrated daily with standard fluorescent microbeads conjugated with definite numbers of FITC molecules (Flow Cytometry Standards Corporation, Research Triangle Park, North Carolina).[32]

Lymphocyte Immunophenotyping

Three to six blood samples of endurance-trained cyclists of the regional and national class were taken according to a standardized procedure at physical rest[28] during normal training and regenerative phases over a period of about 20 months. In addition, each athlete was examined when complaining of typical overload-induced overtraining symptoms such as decrease of performance, soreness of muscles, unusual constant fatigue, disturbances during sleep, and gastrointestinal or other symptoms.[27] Leukocytes in fresh EDTA (ethylenediaminetetraacetic acid)–anticoagulated full blood (20 μL) were incubated at room temperature (RT) for 25 min with directly FITC (fluorescein isothiocyanate)– or PE (phycoerythrin)–labeled antibodies (20 μL) against lymphocyte membrane antigens in combinations of CD45RO/CD4, CD45RO/CD8, CD3/HLA-DR, CD3/CD16, and CD19 and with appropriate IgG1/IgG2 controls (BD and DAKO, Hamburg, Germany). The antibodies were used in 1/2 to 1/5 dilution after appropriate titration experiments in the FACSCAN flow cytometer. Following the antibody incubation, erythrocytes were lysed by FACS lysing solution (BD) for 7 min and centrifuged for 5 min at 250*g* at RT. The cell pellet was shortly vortexed with 2 mL of cell wash (BD), followed by 5 min of 250*g* centrifugation at RT and resuspension of the cell pellet in 0.5 mL Ultracount (BD).

Linear FSC and SSC scatter signals in combination with four-decade logarithmic FITC and PE fluorescence signals of lymphocytes, monocytes, and granulocytes were collected with a FACSCAN (BD) flow cytometer and were stored as FCS1.0 list mode files.[25,26,29,30]

The day-to-day reproducibility for both flow cytometers was assured by a mixture of monosized, fluorescent CaliBRITE (BD) beads of various fluorescence intensity.

Granulocyte and Monocyte Functions

Heparinized fresh venous blood was overlaid onto a cushion of Histopaque-1077 separation medium (Sigma) for 30 min at room temperature to deplete erythrocytes by aggregation at the Histopaque/blood interface and 1*g* sedimentation. The upper two-thirds of the supernatant containing between 0.4×10^7 and 1.2×10^7 leukocytes/mL was carefully removed to avoid any contact of leukocytes with the separating medium and was subsequently stored on ice.[33]

The oxidative activity of granulocytes and monocytes was determined by incubation for 15 min at 37 °C of 1/50 HBS-buffer (0.15 M NaCl, 5 mM HEPES, pH 7.35)–diluted leukocyte supernatant in the presence of 43 μM dihydrorhodamine123 (DHR) (Molecular Probes, Eugene, Oregon)[34] and 60 μmoles of the DNA dye propidium iodide (PI, Sigma) to counterstain the dead cells. DHR-loaded cells were incubated for 15 min at 37 °C either with 10 ng/mL of recombinant tumor necrosis factor–α (TNF-α) (Sigma, produced in yeast, specific activity of 2×10^7 U/mg protein) or with 10^{-7} moles of the bacterial peptide *N*-formyl-Meth-Leu-Phe [FMLP, Sigma, stock of 1 mM in dimethylformamide (DMF)] or they were first primed with FMLP for 15 min followed by TNF-α for another 15 min at 37 °C or were incubated in buffer alone to record the spontaneous cellular oxidative activity. Another leukocyte aliquot was incubated with 10^{-7} M phorbol 12-myristate 13-acetate (PMA, Sigma, stock of 1 mM in DMF) as a positive control.

Intracellular calcium levels were determined following 15 min of 22 °C incubation of leukocytes in HBS-buffer as described earlier in the presence of 20 μM INDO1/AM calcium indicator dye[35] and 60 μM PI. Fifty μL of heparinized (20 U/mL) blood was incubated for 30 min at 37 °C with 5 μL of a suspension of *E. coli* K12 strain (Sigma) at a concentration of 7×10^9 bacteria/mL[36] or with buffer as a control prior to staining with INDO1/AM.

Protease activity measurements were performed with the rhodamine110 (R110) cysteine protease substrates, $(Z\text{-Phe-Arg})_2$-R110 and $(Z\text{-Arg-Arg})_2$-R110, and the serine protease substrate, $(Z\text{-Ala-Ala})_2$-R110, at final concentrations of 4 μM in the absence or presence of 10 μM cysteine proteinase inhibitor Z-Phe-Ala-CHN_2 or 1 mM serine proteinase inhibitor diisopropyl-fluorophosphate (DFP).[33,37]

The flow cytometric measurements for DHR and the protease substrates were performed on a FACSCAN flow cytometer (BD) and the INDO1 measurements were performed on either a PASIII (PARTEC, Münster, Germany) or a FLUVOMETRICELL flow cytometer.[38]

Data Processing

The CLASSIF1 program system (FIGURE 1) operates in personal computers under MS-DOS 3.0 to 5.0 or WINDOWS3.1. The FCS1.0 or FCS2.0 list mode files of up to eight parameters from the Becton Dickinson, Coulter, or PARTEC instruments are processed by the DATLYS procedure, which generates three-, two-, and one-parameter histograms. The two-parameter histograms are obtained through gated or ungated projection of the multidimensional list mode on any wanted coordinate plane and are evaluated by the multiwindow calculation procedure CALC.

Procedure CALC determines for each window the cell or particle content, the mean abscissa and ordinate values, and the ratio and standard deviation of these values supplemented by the respective coefficients of variation. The list mode collected from logarithmic amplifiers is relinearized during the calculation either by a software lookup table for each amplifier generated via standardized internal electronic pulse calibration with a deviation of $<1\%$ (PARTEC) or according to the logarithmic decade information of the manufacturer (Becton Dickinson/Coulter).

Calculated results are introduced by task-specific procedures into binary data bases. A data base decompression (DECOMP) and compression (COMPRESS) function assures either the export of CLASSIF1 data bases in ASCII format for further treatment in external evaluation procedures, for example, dBase or EXCEL program systems, or the import of externally generated ASCII data bases.

Procedure LEARN, in a first step, determines for each data base column the 0, 1, 2, 5, 10, 15, 20, 25, 30, 70, 75, 80, 85, 90, 95, 98, 99, and 100 percentiles of the value distributions of the normal samples. The learning set contains, in the initial phase, data from approximately 50 normal samples and 10 to 30 samples of each disease state from different patients. Procedure LEARN classifies samples by percentile analysis using the three most significantly different columns of each data base for the distinction between normal and abnormal samples. A sample is called abnormal if one of the three classifiers is outside of the 90, 95, 98, 99, or 100 percentile of the distributions for the normal samples of the learning set. In addition, the means, the

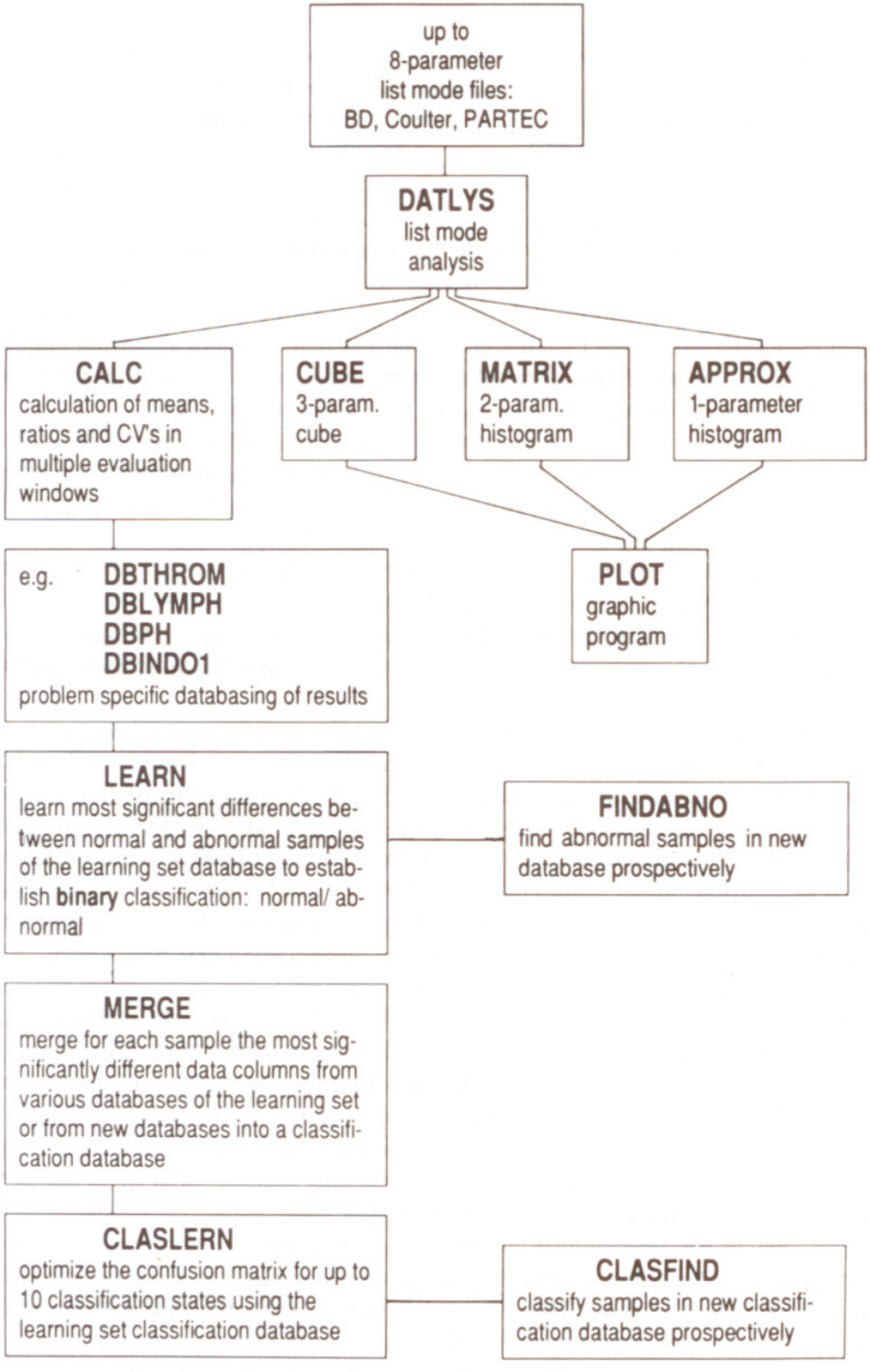

FIGURE 1. Schematic block diagram of the CLASSIF1 program system.

standard deviations (SD), and the standard errors (SE) of each data base column for the normal and abnormal samples of the learning set are determined by procedure LEARN.

Procedure LEARN, in a second step, introduces the 5 data base columns with the most distinctive percentiles into a multifactorial data base after standardization of the values of each data base column onto the respective mean of the normal samples. Multifactors are calculated by multiplying or dividing the imported data base column values of each sample in all possible permutations. Multiplication is performed if the mean of the abnormal samples is higher than the mean of the normal samples and division is used if it is lower. This provides multifactors with maximum discrimination potency. The percentiles of the 27 newly generated data base columns are calculated and the 3 data base columns with the most discriminant percentiles of all data columns are selected for the final distinction between normal and abnormal samples. The percentile values of the selected columns are stored and used by procedure FINDABNO for the prospective classification of unknown samples.

The multifactorial classification in practically all instances is significantly more sensitive than the classification by the original data base columns because several small, but individually nonsignificant differences of single data base columns are concentrated and amplified in appropriate multifactors.

The multifactorial classification is useful for binary decisions, for example, for the automated recognition of cancer cells,[39,40] but it is insufficient for the distinction between several classification states, for example, danger of sepsis, danger of shock, and transitional and stable state in intensive-care patients using functional granulocyte parameters.[31]

Multiple classifications are obtained with the CLASSIF1 program system in the following way: Between 1 and 5 of the percentilewise most distinctive data base columns for each disease state from procedure LEARN from up to 10 different data bases with a maximum of 50 columns per data base are merged by procedure MERGE into a classification data base. Procedure CLASLERN uses a symmetric pair of percentiles, for example, the 10 and 90 or the 15 and 85 percentiles, of the value distribution of the normal samples of each transferred data base column and this generates a triple-state replica matrix of the classification data base where "+" is assigned to column values above the upper percentile, "0" is assigned to values between both percentiles, and "−" is assigned to values below the lower percentile. By repeated iterations, a triple-matrix classification mask for each disease state is optimized from the learning set in such a way that the highest overall distinction is obtained for each disease state in the confusion matrix between clinical diagnosis and the CLASSIF1 classification of the learning set. Data base columns that do not improve the result are omitted during the iterations because their presence in the classification mask deteriorates the overall classification result.

The sum of the percentages in the horizontal lines of the confusion matrices (FIGURE 2) may be higher than 100% when some samples are classified into more than one classification state. This does not, however, affect the values of the correct classifications in the boxed diagonal of the confusion matrices.

The optimized reference classification masks for each disease state are stored for later use by procedure CLASFIND, which prospectively classifies unknown samples. Procedure CLASFIND determines the classification mask for unknown samples by

CONFUSION MATRIX (% PAT)
DATABASE: THC1MR3U.BI4

CLIN. DIAG.	PAT. n	FLOW-CLASSIF. N	R	D
N	17	100.0	.0	.0
R	97	1.0	95.9	3.1
D	14	50.0	7.1	57.1

N =NORMAL THROMBOS
R =INF.RISK THROMBOS
D =DIABETES THROMBOS
10-90% percentil.of normals

CONFUSION MATRIX (% PAT)
DATABASE: GAC1MRG3.BI4

CLIN. DIAG.	PAT. n	FLOW-CLASSIF. N	O
N	51	96.1	11.8
O	21	14.3	95.2

N =NORMAL
O =OVERTRAINING
10-90% percentil.of normals

CONFUSION MATRIX (% PAT)
DATABASE: KEL4.BI4

CLIN. DIAG.	PAT. n	FLOW-CLASSIF. N	O	1	2
N	6	100.0	.0	.0	.0
O	35	8.6	60.0	37.1	2.9
1	27	.0	7.4	88.9	3.7
2	14	.0	14.3	21.4	71.4

N =NORMAL GRANULOCYTES
O =NORM.INT.CARE GRAN.
1 =INFECTION
2 =SEPSIS
15-85% percentil.of normals

FIGURE 2. Confusion matrices (see text).

calculation of their triple matrix according to the percentiles of the optimized reference mask. During the subsequent comparison of the resulting sample mask with the optimized reference mask for each disease state, a sample is classified into the disease state with the highest positionwise coincidence with any one of the reference masks. In case of equally frequent coincidences for several reference masks, the program provides several classifications.

RESULTS

Thrombocytes

Thrombocyte list mode files were evaluated with one FSC/SSC window (FIGURE 3A) and three evaluation windows in the FSC/FITC antibody graph (FIGURE 3B). Window 1 comprised antibody negative and positive thrombocytes, window 2 comprised only the antibody negative thrombocytes, and window 3 comprised only the antibody positive thrombocytes. The boundary between windows 2 and 3 was the same for all list mode files and was set to between 5% and 10% of the maximum thrombocyte frequency of the thrombocyte cluster of normal patients. The evaluation procedure assured that small percentages of antibody positive thrombocytes as well as small increases of antibody fluorescence of all thrombocytes were reliably detected. Eleven values, that is, % antibody positive cells and negative cells, mean FSC, ratio SSC/FSC, mean antibody fluorescence, and mean antibody density of all cells, of the antibody positive and negative cells were extracted from each measurement and were data-based.

After the determination of the 10% and 90% values of the normal values of each data base column, the columns with the highest identification rate for abnormal samples were "learned" with the LEARN procedure. The three most significantly different columns for each of the four measurements (CD62, CD63, thrombospondin, IgG control) and for each of the three classification states (normal, risk for myocardial infarction, diabetes) were merged by procedure MERGE into the classification data base, which theoretically should contain $3 \times 4 \times 3 = 36$ data columns. The merged data base, however, contained only 23 data columns because duplicate data base columns were not introduced into the merged data base. They do not provide additional information and thus prolong the optimization process. Only 8 of the 23 data columns of the merged data base were finally used for the reference classification masks (FIGURE 4) during the iterative optimization process of the CLASLERN procedure. The classification masks permit the classification of the learning set as shown in the confusion matrix (FIGURE 2, left panel). The classification results are given either in the form of an individual letter (FIGURE 5) or as a listing for scientific purposes (FIGURE 6). The reduction of the learning set to 20, 40, 60, and 80 patients and the use of the remaining 77, 57, 37, and 17 patients as unknown samples for a prospective classification indicates that the classifier is stable from a learning set of 40 patients and upwards (TABLE 1).

An interesting aspect of the CLASSIF1 program system is the possibility of standardizing each column of the classification data base onto the respective mean of the normal patients of the learning set. The optimized classification reference mask can then be expressed in a standardized, that is, instrument-independent, way

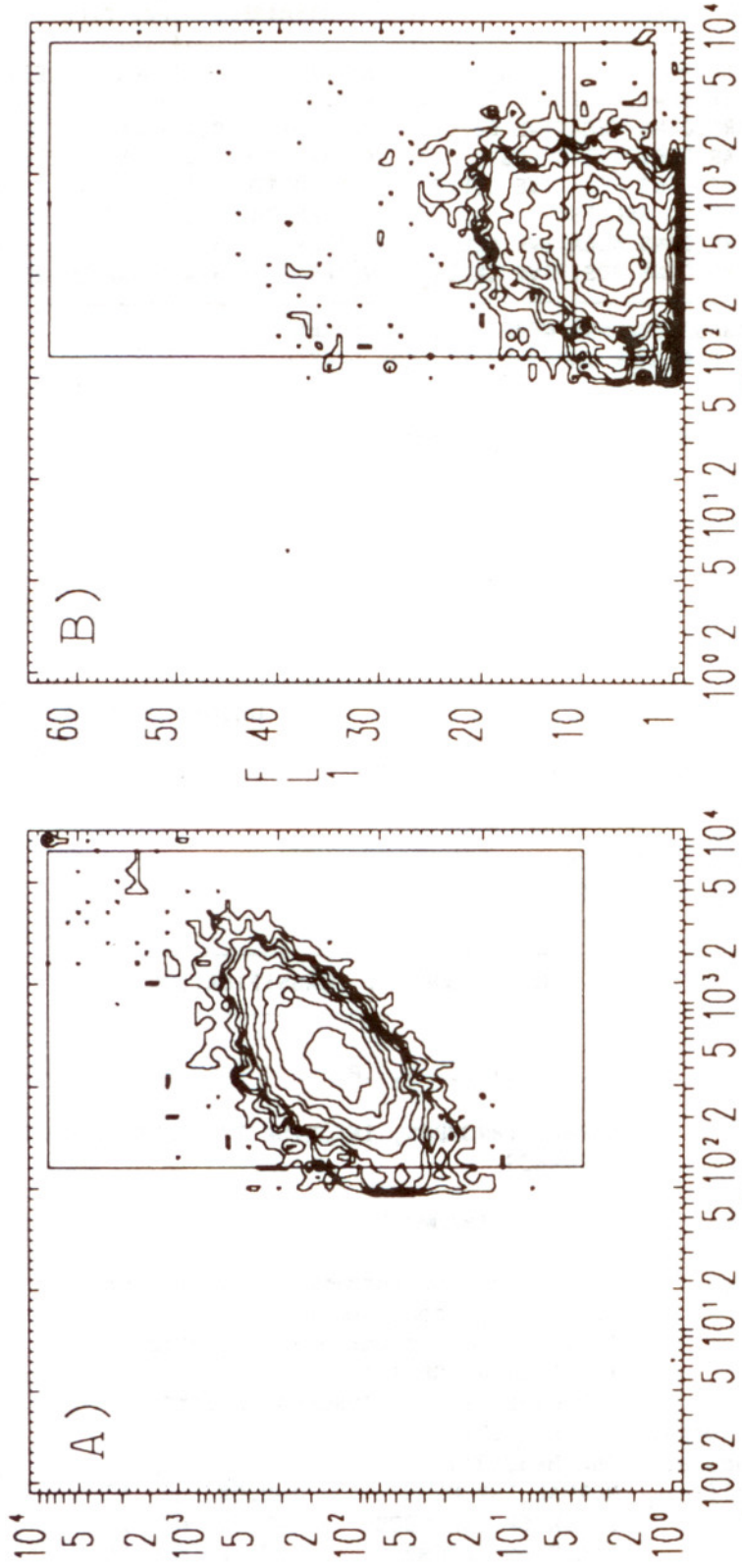

FIGURE 3. Light scatter (A) and fluorescence (B) measurements of anti-CD62 incubated thrombocytes from a patient with high risk for myocardial infarction (TS0051).

LEARNING SET DATABASE: THC1MR3U.BI4

MASK PAR.	DB- COL. PARAMETER EXPLANATION	ORIG COL.	ORIGINAL DATABASE	ORIGINAL REF.FILE	N	R	D
1	2=MEAN AB POSITIVE CELLS	6	THRMIGGU.BI4	THRNIGGU.RE4	0	+	0
2	4=AB SURF.DENS.POSITIVE CELLS	9	THRMIGGU.BI4	THRNIGGU.RE4	0	+	0
3	5=AB SURF.DENS.NEGATIVE CELLS	10	THRMIGGU.BI4	THRNIGGU.RE4	0	0	-
4	8=MEAN AB POSITIVE CELLS	6	THRM217U.BI4	THRN217U.RE4	0	+	0
5	10=AB SURF.DENS.POSITIVE CELLS	9	THRM217U.BI4	THRN217U.RE4	0	+	0
6	12=MEAN RATIO SSC/FSC	4	THRM228U.BI4	THRN228U.RE4	0	+	-
7	17=AB SURF.DENS.NEGATIVE CELLS	10	THRM228U.BI4	THRN228U.RE4	0	-	-
8	22=AB SURF.DENS.POSITIVE CELLS	9	THRMP10U.BI4	THRNP10U.RE4	0	+	0

8 OF 23 DATABASE COLUMNS WERE USED FOR MASKS
74 ITERATIONS FOR MASK IMPROVEMENT WERE PERFORMED

FIGURE 4. The merged learning set data base (see text).

Max-Planck-Institut
für Biochemie
Arb.Gruppe Zellbiochemie
Am Klopferspitz
8033 Martinsried
Tel: 49/89/8578-2518, -2525
Fax: 49/89/8678-2563, -3777

Date: Oct. 2, 1992
Time: 15:11:48h

AUTOMATED FLOW CYTOMETRIC CLASSIFICATION

Sample: ?...00224.....R.

QUESTION: NORMAL THROMBOS, INF.RISK THROMBOS, DIABETES THROMBOS

CLASSIFICATION: INF.RISK THROMBOS

Measurement: DIII assay for thrombocyte antigens:
CD62,CD63,thrombospondin +
forward and sideward light scatter
!!! TEST PHASE !!!
- PLEASE, REPEAT ABNORMAL ASSAYS

Classific.database: THC1M3UT.BI6
Learning database: THC1MR3U.BI4
Correct classific: NORMAL THROMBOS : 100.0% n= 17
INF.RISK THROMBOS : 95.9% n= 97
DIABETES THROMBOS : 57.1% n= 14

Signature: Prof.Dr.G.Valet

FIGURE 5. Classification results given in the form of an individual letter (see text).

NR.	CLINICAL DIAGNOSIS	ABBREVIATION	COIN	CLASSIFICATION MASK
1	NORMAL THROMBOS	N	1.00	00000000
2	INF.RISK THROMBOS	R	1.00	++0+++-+
3	DIABETES THROMBOS	D	1.00	00-00--0

REC. NR.	DATAB: THC1MR3U.BI4 DATAB.RECORD LABEL	FLOW-CLASSIFICATION	CLAS COIN	CLASSIFIC.INDICATORS
1	N...00101..... .	N	.88	000000-0
2	N...00102..... .	N	1.00	00000000
3	N...00103..... .	N	.63	+00++000
4	N...00104..... .	N	.63	000++0-0
5	N...00105..... .	N	.88	000000+0
6	N...00106..... .	N	.63	0++000+0
7	N...00107..... .	N	.88	00000+00
8	N...00108..... .	N	.88	+0000000
9	N...00109..... .	N	.50	0--0-00-
10	N...00110..... .	N	.88	0000000+
11	N...00111..... .	N	.75	00+0000+
12	N...00112..... .	N	1.00	00000000
13	N...00113..... .	N	1.00	00000000
14	N...00114..... .	N	1.00	00000000
15	N...00115..... .	N	.88	00000-00
16	N...00116..... .	N	.75	-00-0000
17	N...00117..... .	N	.63	++000+00
36	R...00136..... .	R	.63	++0++000
37	R...00137..... .	R	.88	++0+++0+
38	R...00138..... .	R	.63	+0-++0-+
39	R...00139..... .	R	.75	++-++0-+
40	R...00140..... .	R	.75	++++++0+
41	R...00141..... .	R	.75	++++++0+
42	R...00142..... .	R	.75	++++++++
43	R...00143..... .	R	.75	++-++0-+
44	R...00144..... .	R	.88	++0++0-+
45	R...00145..... .	R	.63	++-+0--+
46	R...00146..... .	R	.75	+00++--+
47	R...00147..... .	R	.75	++-++0-+
48	R...00148..... .	R	.88	++0++0-+
49	R...00149..... .	R	.75	++0++0-.
50	R...00150..... .	R	.75	+00++0-+
51	R...00151..... .	R	.63	+0-++0-+
52	R...00152..... .	R	.75	+00++0-+

FIGURE 6. Classification results given as a listing (see text).

(TABLE 2), which is an important prerequisite for the laboratory-independent standardization of flow cytometric classifiers.

Lymphocytes

The results from five flow cytometric lymphocyte phenotypes and the IgG control were calculated by quadrant evaluation (FIGURE 7) as % positive cells (FIGURE 8A),

TABLE 1. Predictive Value of the Risk of Myocardial Infarction Classification

Patients (n/n)[a]	Correctness of R (%)[b]
77/20	80
57/40	93
37/60	95
17/80	94

[a]Number of predictive patients/number of learning set patients.
[b]R = risk of myocardial infarction.

as relative intensity (FIGURE 8B), and as cell surface density of the antigen expression from the antigen content and the FSC by considering the FSC signal as a cell volume equivalent. The results from each phenotype were introduced into a separate data base. Self-learning on a set of normal and clinically overtrained competition cyclists was then performed. The CLASLERN procedure transferred a total of 25 data base columns from the various lymphocyte phenotype measurements into the classification data base. Only six parameters from the CD45RO/CD4 and CD45RO/CD8 measurements were utilized by the optimized reference classification masks. The confusion matrix (FIGURE 2, middle panel) shows that the overload-induced overtraining syndrome is well recognized. The increased amount (FIGURE 8B) and the relative surface density of CD45RO antigen expression on T lymphocytes, as demonstrated by fluorescent antibody binding, are the most informative parameters.

Granulocytes and Monocytes

Leukocyte function can be estimated by oxidative activity (FIGURES 9A and 9B), by intracellular calcium levels (FIGURES 9C and 9D), and by protease activity (FIGURES 10A–D). The confusion matrix (FIGURE 2, right panel) of a series of DHR leukocyte measurements of healthy adult persons, of complication-free ICU patients, and of ICU patients with infection or sepsis shows that the oxidative activity of

TABLE 2. Standardized Classification Formula for Thrombocytes upon Incubation with Isotype IgG or CD62, CD63, and Thrombospondin Antibodies[a]

Parameters	IgG Mean Ab Pos Cells	IgG Srf Ab Pos Cells	IgG Srf Ab Neg Cells	CD62 Mean Ab Pos Cells	CD62 Srf Ab Pos Cells	CD63 Mean SS/FS Ratio	CD63 Srf Ab Neg Cells	Thrmsp Srf Ab Pos Cells
Upper threshold	1.03	1.12	1.09	1.09	1.12	1.12	1.09	1.10
Lower threshold	0.96	0.88	0.88	0.94	0.86	0.88	0.85	0.86
Normals	0	0	0	0	0	0	0	0
Risk myocard. inf.	+	+	0	+	+	+	−	+
Diabetes	0	0	−	0	0	−	−	0

[a]Standardized thresholds: 10% and 90% of normals, expressed as the ratio of percentile/mean. Terms: Thrmsp = thrombospondin, Ab = antibody, Srf = relative surface density, SS = 90°, FS = forward light scatter, Pos = positive, Neg = negative.

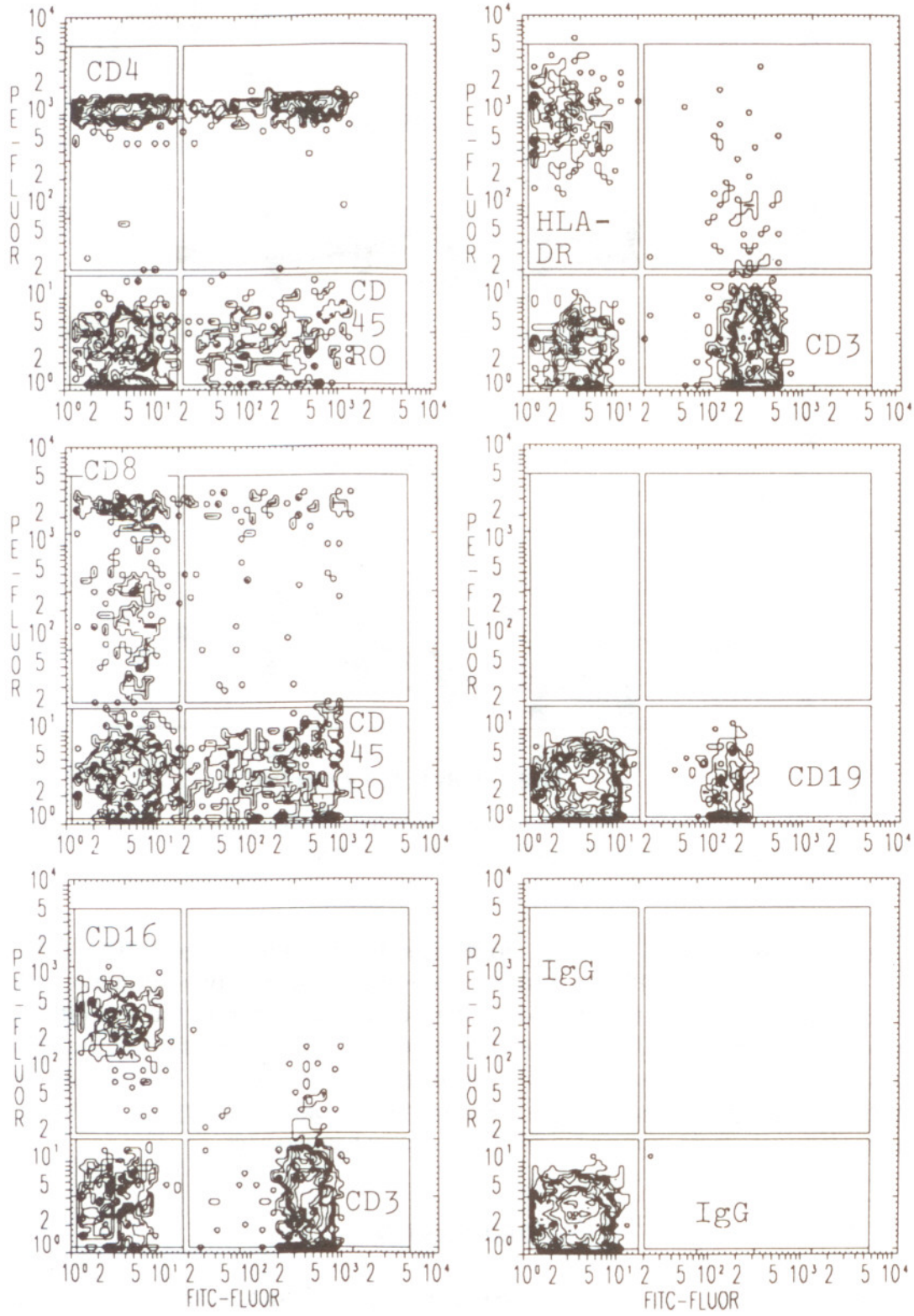

FIGURE 7. Lymphocyte immunophenotypes in the peripheral blood of an endurance-trained competition cyclist with overtraining syndrome (GA011).

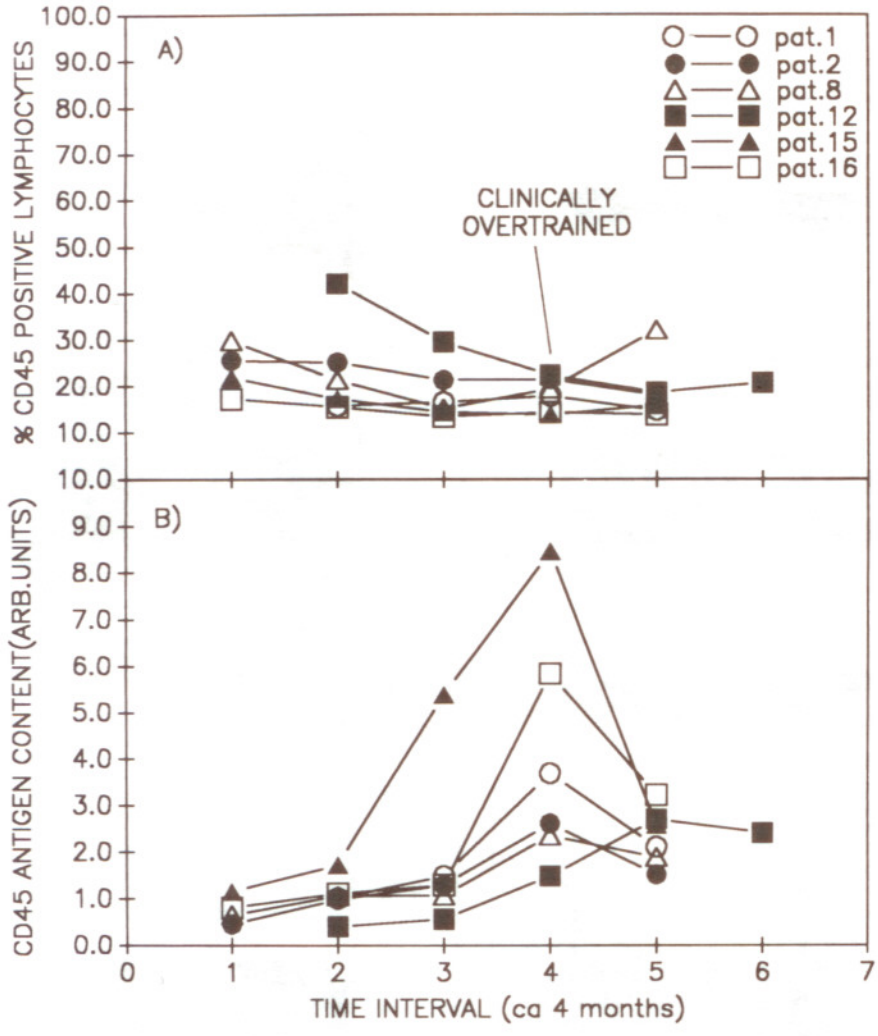

FIGURE 8. Percent CD45RO positive (A) and intensity of antigen expression (B) on peripheral blood lymphocytes of endurance-trained competition cyclists.

granulocytes and monocytes either alone or under the influence of TNF-α, FMLP, FMLP + TNF-α, and PMA contains a significant degree of information. The CLASLERN procedure transferred 42 data base columns from the five different measurements into the classification data base. The optimized reference classification masks contain 32 parameters from all five measurements. Interestingly and unexpectedly, only 17 parameters were selected from granulocytes, whereas 13 parameters were from monocytes and 2 were from lymphocytes.

DISCUSSION

The results show that the CLASSIF1 program system classifies the individual patient status with a high degree of accuracy from list mode data, although the evaluation of the same data by the software of the various flow cytometers was inconclusive at the individual patient level.

The actual performance of the CLASSIF1 program system depends on the speed of the computer and on the complexity and multiplicity of the list mode data, but it

remained in all of the aforementioned evaluated cases within 5 to 10 min of learning time for learning sets on the order of 100 to 200 patients derived from 500 to 1000 list mode measurements. In contrast, the prospective classification of unknown samples proceeds at a much higher speed of approximately 100 decisions/s in a computer with an 80486/487 processor at 33-MHz clock frequency. The different list mode calculations that have to be performed by any evaluation program take an additional 50 to 60 s per patient; that is, the CLASSIF1 program can be used in an automatically operating flow cytometer when the list mode data are rapidly (60–80 kbytes/s) transferred, for example, via ethernet communication from the flow cytometer to the evaluation computer and calculated while the flow cytometer measures the next sample.

The establishment of instrument-independent classifiers opens the way to the

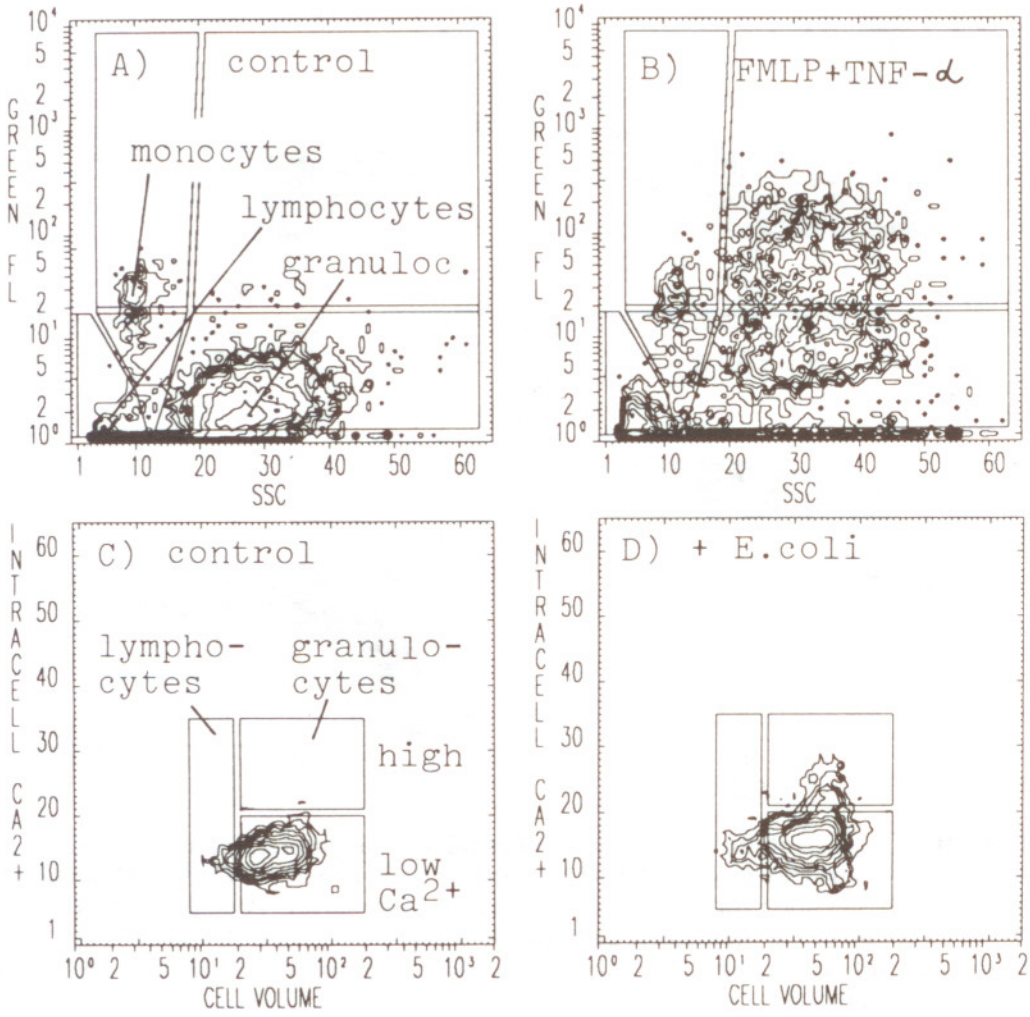

FIGURE 9. Dihydrorhodamine123 (DHR) oxidation in unstimulated (A) and 10^{-7} M FMLP + 10 ng/mL TNF-α stimulated (B) peripheral blood leukocytes of an ICU patient (KE1014) whose monocytes are spontaneously stimulated. The control assay was 0.46% DHR positive (i.e., stimulated granulocytes), TNF-α alone caused an increase to 5.79%, FMLP alone caused an increase to 11.7%, FMLP + TNF-α caused an increase to 41.4%, and 10^{-7} M PMA alone caused an increase to 89.9%. The granulocytes of an infected ICU patient (KE0074) show normal Ca^{2+} levels when incubated in the presence of 20 μM INDO1/AM ester (C), whereas the Ca^{2+} level cannot be restored following 30 min of 37 °C incubation of heparinized blood with *E. coli* K12 bacteria (D).

standardization of flow cytometric classifiers. The standardization depends on a flow cytometer's potential to reproduce the typical histogram patterns for a given application, on the long-term intralaboratory standardization of the measurements by fluorescent particles, and on reproducibly produced antibodies or biochemical reagents. Expert groups can exchange and classify their standardized data bases to obtain a consensus on the classifier formula (TABLE 2) for a given application. At a

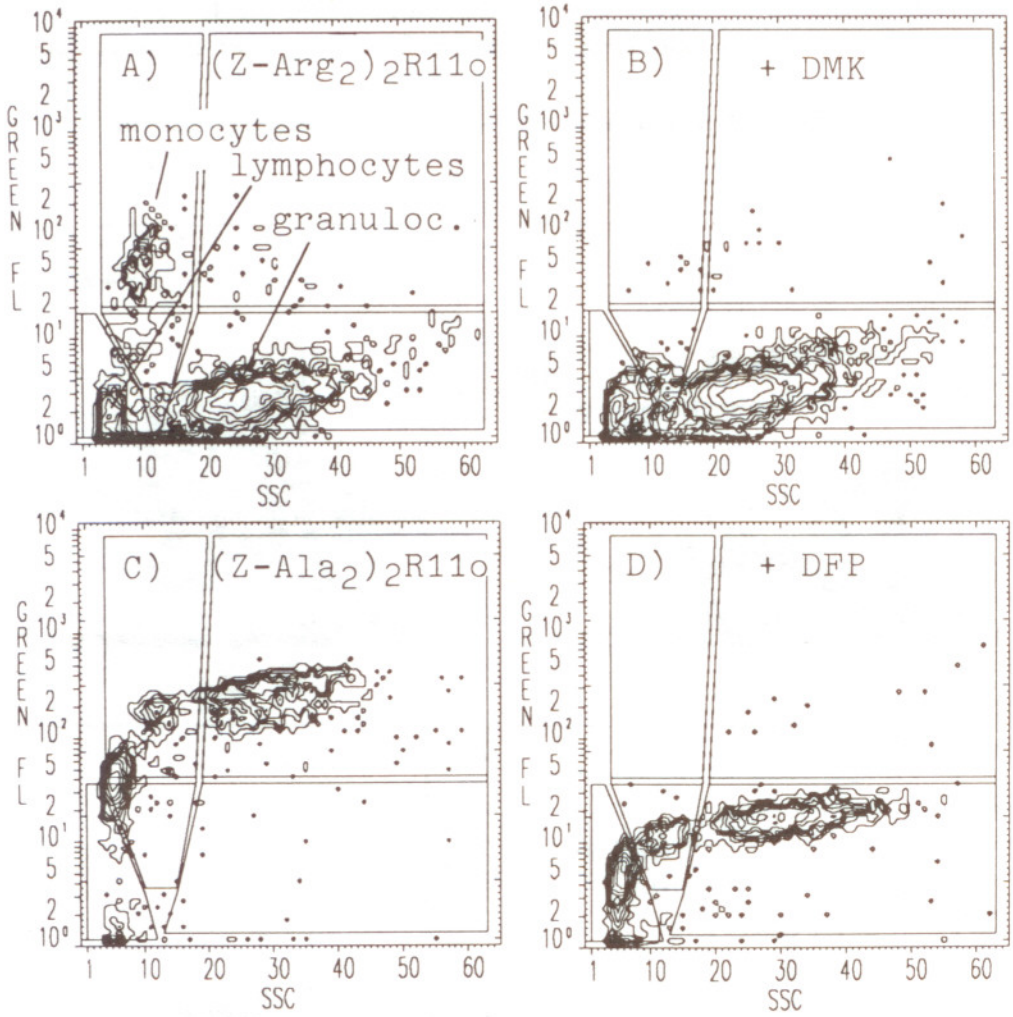

FIGURE 10. Cleavage of the fluorogenic cathepsin L cysteine proteinase substrate $(Z\text{-}Arg_2)_2$-R110 (4 μM) by peripheral human blood leukocytes for 20 min at 37 °C (A) and inhibition of cleavage by preincubation for 10 min at 37 °C with the inhibitor Z-Phe-Ala-CHN_2 (DMK, 10 μM) (B). Cleavage of the fluorogenic cathepsin G serine proteinase substrate $(Z\text{-}Ala_2)_2$-R110 at the same conditions as above (C) and preincubation of the assay with inhibitor DFP (1 mM) as a specificity control (D).

new location, only the necessary parameter combinations of a group of normal individuals have to be measured. The standardized data of this patient group are then cross-classified against the expert consensus group's data set of normal individuals in order to certify conformity with the consensus data set. However, the establishment of additional learning sets for abnormal individuals at the new

location is unnecessary. This feature of the CLASSIF1 classifiers seems of high practical importance.

The proposed classification concept does not require an absolute numerical coincidence of the unstandardized individual parameters between different flow cytometers in various laboratories because the classification does not depend on individual parameter values, but on the relative changes of the parameter patterns. Nevertheless, it is clear that standardization trials should bring the numerical values of the individual parameters of flow cytometers in different laboratories as close as possible.

The aforementioned classifications (FIGURE 2) show a close correlation between the biochemical and immunological parameter changes and disease states. Due to the fact that diseases are caused by biochemical deviations in cells and cellular systems, the findings are quite compatible with this general concept. All investigated cell types participate to some extent directly in the disease process, for example, the thrombocytes (FIGURE 2, left panel) in the potential occlusion of narrowed or altered blood vessels preceding the acute event of myocardial infarction or the development of diabetic angiopathy, the lymphocytes (FIGURE 2, middle panel) in the altered defense against infection during an overload-induced overtraining syndrome of competition cyclists, and the granulocytes (FIGURE 2, right panel) of intensive-care-unit patients due to their altered cell functions.

The unexpected use of certain data base columns for the reference emphasizes the need for the invasive self-learning data evaluation of the CLASSIF1 program system, which does not depend on a priori knowledge of the complex parameter interdependences; in fact, this knowledge usually does not exist. Note that this frequent lack of a priori expert knowledge on the interpretation of complex flow cytometric measurements constitutes a significant problem for the use of expert systems[14–17] in flow cytometry, but can be avoided by employing self-learning expert systems.

The execution speed of the triple-matrix classification program is rapid because learning and classification are performed on comparatively few numbers from preestablished data bases. The additional list mode calculations for the data bases of the CLASSIF1 program system are simple and suitable for parallel processing, that is, the data from a sample just measured in an automated flow cytometer can be off-line classified in a second processor during the measurement time of the next sample, leading to a quasi on-line performance of this classification methodology.

In contrast, cluster algorithms,[1,2] principal component/biplot analysis,[9,10] and neural networks[12] are capable of direct on-line list mode classification. This requires substantially more calculations for the classification as compared with the CLASSIF1 program system, but offers the principal advantage of classified on-line cell sorting or of on-line sample classification in automated flow cytometers. However, because efficient classifications require mostly the consideration of data from several multi-parameter measurements, a one-step on-line list mode classification is usually not achievable.

Therefore, the on-line classification capacity of a classification method may not represent in practice a decisive advantage because the measurement time rather than the time of computation is the speed-limiting process for the operation of automated flow cytometers. In this situation, some of the CLASSIF1 features like

simplicity, easy adaptability to new problems, and generation of instrument-independent and standardized classifiers seem of high practical interest.

REFERENCES

1. BIERRE, P., R. MICKAELS & D. THIEL. 1991. Multidimensional visualization and autoclustering of flow cytometric data. Cytometry **12**(suppl. 5): 64.
2. SALZMAN, G. C., S. R. MCLAUGHLIN, S. P. ELLIS & J. F. LEARY. 1991. 3-D autostereoscopic viewing of multidimensional data for guided cluster analysis. Cytometry **12**(suppl. 5): 64.
3. MURPHY, R. F. 1985. Automated identification of subpopulations in flow cytometric list mode data using cluster analysis. Cytometry **6:** 302–309.
4. BECKMAN, R. J. 1988. Warning: your cluster algorithm may be dangerous to your analysis. Cytometry **9**(suppl. 2): 6.
5. SALZMAN, G. C., S. P. PEDERSON, R. J. BECKMAN, R. B. KRALL & C. C. STEWART. 1988. Cluster analysis for flow cytometric data. Cytometry **9**(suppl. 2): 84.
6. DEMERS, S., J. KIM, P. LEGENDRE & L. LEGENDRE. 1992. Analyzing multivariate flow cytometric data in aquatic sciences. Cytometry **13:** 291–298.
7. VALET, G., S. BAMBERGER, H. HOFMANN, R. SCHINDLER & G. RUHENSTROTH-BAUER. 1979. Flow cytometry as a new method for the measurement of electrophoretic mobility of erythrocytes using membrane charge staining by fluoresceinated polycations. J. Histochem. Cytochem. **27:** 342–349.
8. VALET, G. 1990. Graphical representation of three-parameter flow cytometer histograms by a newly developed FORTRAN IV computer program. *In* Flow Cytometry IV. O. D. Laerum, T. Lindmo & E. Thorud, Eds.: 125–129. Universitetsferlaget. Oslo.
9. LEARY, J. F., S. ELLIS, S. HESPELT, S. R. MCLAUGHLIN & J. G. GRAM. 1988. Principal component/biplot analysis for correct identification of rare human fetal cells in maternal blood. Cytometry **9**(suppl. 2): 29.
10. LEARY, J. F., S. R. MCLAUGHLIN, M. A. CORIO, L. REMLEY, J. G. GRAM & S. BURDE. 1990. High speed principal component/biplot sorting of rare cells. Cytometry **11**(suppl. 4): 19.
11. FRANKEL, D. S., R. J. OLSON, S. L. FRANKEL & S. W. CHISHOLM. 1989. Use of a neural net computer system for analysis of flow cytometric data of phytoplankton populations. Cytometry **10:** 540–550.
12. FRANKEL, D. S. & S. L. FRANKEL. 1991. Real time neural network for flow cytometry analysis. Cytometry **12**(suppl. 5): 63.
13. REDELMAN, D. 1991. Neural network analysis of flow cytometry data. Cytometry **12**(suppl. 5): 63.
14. SALZMAN, G. 1987. A knowledge-based system as a cluster assistant. Cytometry **8**(suppl. 1): 12.
15. BAGWELL, C. B. 1988. New horizons: expert systems for flow cytometry. Cytometry **9**(suppl. 3): 89–93.
16. MCGUIRE, D., J. P. ROBINSON & G. B. KING. 1991. Automated interpretation of phenotypic data using a multiparameter expert system. Cytometry **12**(suppl. 5): 63–64.
17. ROBINSON, J. P., K. RAGHEB, G. LAWLER, S. KELLEY & G. DURACK. 1992. Rapid multivariate analysis and display of cross-reacting antibodies on human leukocytes. Cytometry **13:** 75–82.
18. VALET, G., H. H. WARNECKE & H. KAHLE. 1987. Automated diagnosis of malignant and other abnormal cells by flow-cytometry using the DIAGNOS1 program system. *In* Clinical Cytometry and Histometry. G. Burger, J. S. Ploem & K. Goerttler, Eds.: 58–65. Academic Press. New York/London.
19. HABBERSETT, M. C., M. SHAPIRO, B. BUNNAG, I. NISHIYA & CH. HERMAN. 1979. Quantitative analysis of flow microfluorometric data for screening gynecologic cytology samples. J. Histochem. Cytochem. **27:** 536–544.
20. TSCHÖPE, D. & B. SCHWIPPERT. 1992. Düsseldorf III Assay: Bestimmung von Thrombozytenmarkern auf Thrombozyten im Durchflußzytometer. *In* Klinische Zytometrie. G. Schmitz, Ed. Schattauer. Stuttgart. In press.

21. TSCHÖPE, D., P. RÖSEN, B. SCHWIPPERT & F. A. GRIES. 1992. Platelets in diabetes: their role in the hemostatic regulation in atherosclerosis. Semin. Thromb. Hemostas. In press.
22. TSCHÖPE, D., H. P. SCHULTHEISS, P. KOLAROV, H. K. NIEUWENHUIS, K. DANEHL, B. STRAUER & F. A. GRIES. 1991. Platelet activation is predictive for an increased PTCA-risk. Circulation **84**(suppl. 4): 690.
23. TSCHÖPE, D., J. ESSER, B. SCHWIPPERT, P. RÖSEN, B. KEHREL, H. K. NIEUWENHUIS & F. A. GRIES. 1991. Large platelets circulate in an activated state in diabetes mellitus. Semin. Thromb. Hemostas. **17:** 433–439.
24. TSCHÖPE, D., P. SPANGENBERG, J. ESSER, B. SCHWIPPERT, B. KEHREL, P. RÖSEN & F. A. GRIES. 1989. Flow cytometric detection of surface membrane alterations and concomitant changes in the cytoskeletal actin status of activated platelets. Cytometry **11:** 652–656.
25. GABRIEL, H., L. SCHWARZ, P. BORN & W. KINDERMANN. 1992. Differential mobilization of leucocyte and lymphocyte subpopulations into circulation during endurance exercise. Eur. J. Appl. Physiol. In press.
26. GABRIEL, H., L. SCHWARZ, W. STEFFENS & W. KINDERMANN. 1992. Immunoregulatory hormones, circulating leucocytes, and lymphocyte subpopulations before and after endurance exercise of different intensities. Int. J. Sports Med. **13:** 359–366.
27. FRY, R. W., A. R. MORTON & D. KEAST. 1991. Overtraining in athletes—an update. Sports Med. **12:** 32–65.
28. WESTERMANN, J. & R. PABST. 1990. Lymphocyte subsets in the blood: a diagnostic window on the lymphoid system? Immunol. Today **11:** 406–410.
29. GABRIEL, H., A. URHAUSEN & W. KINDERMANN. 1991. Circulating leucocyte and lymphocyte subpopulations before and after intensive endurance exercise to exhaustion. Eur. J. Appl. Physiol. **63:** 449–457.
30. GABRIEL, H., A. URHAUSEN & W. KINDERMANN. 1992. Mobilization of circulating leucocyte and lymphocyte subpopulations before and after intensive endurance exercise to exhaustion. Eur. J. Appl. Physiol. **65:** 164–170.
31. ROTHE, G., W. KELLERMANN & G. VALET. 1990. Flow cytometric parameters of neutrophil function as early indicators of sepsis- or trauma-related pulmonary or cardiovascular organ failure. J. Lab. Clin. Med. **115:** 52–61.
32. TSCHÖPE, D., P. RÖSEN, B. SCHWIPPERT, B. KEHREL, S. SCHAUSEIL, J. ESSER & F. A. GRIES. 1990. Platelet analysis using flow cytometric procedures. Platelets **1:** 127–133.
33. ROTHE, G., S. KLINGEL, I. ASSFALG-MACHLEIDT, W. MACHLEIDT, CH. ZIRKELBACH, R. B. BANATI, W. F. MANGEL & G. VALET. 1992. Flow cytometric analysis of protease activities in vital cells. Biol. Chem. Hoppe Seyler **373:** 547–554.
34. ROTHE, G., A. OSER & G. VALET. 1988. Dihydrorhodamine123: a new flow cytometric indicator for respiratory burst activity in neutrophil granulocytes. Naturwissenschaften **75:** 354–355.
35. VALET, G., A. RAFFAEL & L. RÜSSMANN. 1985. Determination of intracellular calcium in vital cells by flow cytometry. Naturwissenschaften **72:** 600–602.
36. ROTHE, G. & G. VALET. 1988. Phagocytosis, intracellular pH, and cell volume in the multifunctional analysis of granulocytes by flow cytometry. Cytometry **9:** 316–324.
37. ASSFALG-MACHLEIDT, I., G. ROTHE, S. KLINGEL, R. NATI, W. F. MANGEL, G. VALET & W. MACHLEIDT. 1992. Membrane permeable fluorogenic rhodamine substrates for selective determination of cathepsin L. Biol. Chem. Hoppe Seyler **373:** 433–440.
38. KACHEL, V., E. GLOSSNER, E. KORDWIG & G. RUHENSTROTH-BAUER. 1977. FLUVOMETRICELL, a combined cell volume and cell fluorescence analyzer. J. Histochem. Cytochem. **25:** 804–812.
39. LIEWALD, F., N. DEMMEL, R. WIRSCHING, H. KAHLE & G. VALET. 1990. Intracellular pH, esterase activity, and DNA measurements of human lung carcinomas by flow cytometry. Cytometry **11:** 341–348.
40. VALET, G., L. RÜSSMANN & R. WIRSCHING. 1984. Automated flow-cytometric identification of colo-rectal tumour cells by simultaneous DNA, CEA-antibody, and cell volume measurements. Clin. Chem. Clin. Biochem. **22:** 935–942.