

Automated identification and characterisation of microbial populations using flow cytometry: the AIMS project*

RICHARD JONKER¹, RENÉ GROBEN², GLEN TARRAN³, LINDA MEDLIN², MALCOLM WILKINS⁴, LAURA GARCIA⁵, LAURA ZABALA⁵ and LYNNE BODDY⁴

¹AquaSense, Kruislaan 411A, 1090 HC Amsterdam, The Netherlands.

Fax: +31 20 5922249. E-mail: rjonker@aquasense.com

²Alfred Wegener Institute for Polar and Marine Research, Am Handelshafen 12, 27570 Bremerhaven, Germany.

³Centre for Coastal and Marine Sciences, Plymouth Marine Laboratory, Prospect Place, West Hoe, Plymouth, Devon PL1 3DH, United Kingdom.

⁴Cardiff School of Biosciences, University of Wales, Main Building, Park Place, Cardiff CF1 3TL, Wales.

⁵University of Malaga, Campus de Teatinos, 29071 Malaga, Spain.

SUMMARY: The AIMS (Automatic Identification and characterisation of Microbial populationS) project is developing and integrating flow cytometric technology for the identification of microbial cell populations and the determination of their cellular characteristics. This involves applying neural network approaches and molecular probes to the identification of cell populations, and deriving and verifying algorithms for assessing the chemical, optical and morphometric characteristics of these populations. The products of AIMS will be calibrated data, protocols, algorithms and software designed to turn flow cytometric observations into a data matrix of the abundance and cellular characteristics of identifiable populations. This paper describes the general approach of the AIMS project, with details on the application of artificial neural nets and rRNA oligonucleotide probes.

Key words: plankton, phytoplankton, flow cytometry, artificial neural networks, rRNA probes.

INTRODUCTION

Understanding the population dynamics of planktonic microbial communities is central to many contemporary concerns in marine science. Such concerns include fuelling the food webs that lead to fisheries, assessing the responses of coastal waters to pollution and determining the role of the ocean in sequestering anthropogenically produced carbon dioxide. Marine microbial communities are taxonomically and functionally diverse, comprising

phyto- and zooplankton, bacteria and protozoa. Moreover, marine plankton exhibits a wide range of scales of variability, both in space and time. This variability can only be explained through an understanding of the interactions between different biological and physical processes. A prerequisite for obtaining this knowledge is the ability to sample and analyse the physical, chemical and biological properties of marine waters on appropriate time and space scales.

Some biological variables, such as chlorophyll fluorescence, can be measured continuously. However, assessment of changes in the population struc-

*Received June 22, 1999. Accepted February 1, 2000.

ture of marine microbial populations is either very basic or non-descriptive, such as quantifying total chlorophyll in seawater, or detailed but very labour intensive, such as microscopy. Microscopy allows taxonomic analysis of phytoplankton with relatively rigid cell walls (e.g., armoured dinoflagellates, diatoms and coccolithophorids) whose overall cell morphology is maintained when preserved. However, naked groups, such as small (generally less than 10 μm) flagellated phytoplankton, are often rendered unrecognisable when preserved. This, coupled with microscopical resolution limits, makes it difficult, if not impossible to identify and quantify this preserved phytoplankton with any precision even by trained personnel. This is potentially a major problem, particularly as HPLC analysis of phytoplankton pigments in seawater has shown that in many areas of the ocean small flagellated phytoplankton are the dominant groups of primary producers (Barlow *et al.*, 1999; Bidigare and Ondrusek, 1996; Barlow *et al.*, 1993).

Analytical flow cytometry (AFC) is a methodological approach that permits the precise, rapid, repetitive description of the population structure of primary producers, bacterioplankton and zooplankton. AFC obviates many of the problems associated with microscopy and other analytical methods, offering considerable potential for the rapid, accurate and precise analysis of phytoplankton and bacterial populations (Jonker *et al.*, 1995). However, exploiting flow cytometry for examining marine microbial processes is still limited by data analysis techniques. Most of the currently available AFC software provides single parameter histogram distributions and dual parameter scatter plots. However, the optical data often contain 5-11 measurements or parameters that could be used for discrimination. The vast quantities of multivariate data generated by flow cytometers provide a considerable challenge for data analysis. Utilising the multi-parametric data requires complex multi-variate data analysis techniques. Whereas multivariate statistical approaches have been used (Demers *et al.* 1992; Carr *et al.* 1996), and can be very successful if the appropriate technique can be found, this is often not simple, and invalid assumptions about distributions can cause major problems. Artificial neural networks (ANNs) on the other hand, do not require a-priori knowledge of underlying distributions, once trained they can make identifications in near real-time and have been shown to have considerable potential for identifying phytoplankton from AFC data (e.g., Frankel *et al.*,

1989, 1996; Boddy *et al.*, 1994; Wilkins *et al.*, 1994, 1996) and from morphometric data (e.g., Culverhouse *et al.*, 1994, 1996; Williams *et al.*, 1994).

Species identification is not always possible based on light scatter and fluorescence characteristics, due either to similarities in the optical characteristics between species or to certain species having a wide range of optical characteristics, such as with clumped cells or chains. One possible solution to these problems is to develop species-specific oligonucleotide probes that hybridise only to their target species regardless of their morphology or life cycle stage. The flow cytometer or the fluorescence microscope can detect a fluorescent label attached to these probes and the probe-conferred fluorescence identifies the target, thus verifying species identification. Fluorescent rRNA probes also have great potential for discriminating other components of the plankton, particularly bacteria (Amann *et al.*, 1995; Giovannoni *et al.*, 1995) and protozoa (Rice *et al.* 1997a, b), which do not contain autofluorescent pigments. Developing probes for species or wider taxonomic groups is of fundamental importance in providing comprehensive descriptions of microbial communities and for assessing biodiversity at various taxonomic hierarchies.

The measurements of light scatter and fluorescence taken by the flow cytometer characterise the cells and can therefore be used to calculate optical and chemical properties, i.e., cell size or contents. The time of flight of a particle/cell through the measurement beam is related to a length scale of the particle/cell, and the amount of forward and side scattering of light by individual cells is related to shape of the cell in suspension. The red fluorescence arising from the stimulation of chlorophyll a (and associated pigments) as a cell passes through the measurement beam is related to cell chlorophyll content (Burkill and Mantoura, 1990; Hofstraat *et al.*, 1994) and is also related to chlorophyll concentration in field samples as well (Jonker *et al.*, 1995; Li *et al.*, 1993; Veldhuis *et al.*, 1997).

The light scattering and absorption properties of a microbe are directly related to the size, shape and refractive index of the cell (Spinrad and Brown, 1986; Morel, 1991). The refractive index of the cell is determined by its chemical composition (Barer and Joseph, 1954). There are two components of the refractive index of a cell; the real part and the imaginary part. The real part of the refractive index, which is important to cellular light scattering, is related to the intracellular organic carbon concentra-

tion. In phytoplankton, the imaginary part of the refractive index, which is important to light absorption, is related to the intracellular pigment concentration. Using these relationships, algorithms to estimate pigment and carbon content of microbial cells from flow cytometric measurements will be developed, based on the procedures of Stramski *et al.* (1988) and Stramski and Reynolds (1993).

The main objectives of the AIMS project are to (i) provide software to automate identification and characterisation, including artificial neural nets; (ii) provide new data on inherent properties of marine microbes used for algorithm development to compute inherent properties of cells and (iii) develop molecular markers for identification and characterisation of marine plankton. The application of artificial neural networks and the development and application of molecular probes within

the framework of the AIMS approach are addressed in this paper.

APPLICATION OF ARTIFICIAL NEURAL NETWORKS (ANNs)

Radial Basis Function (RBF) ANN are being used as their identification success is at least as good as other ANN paradigms and non-neural classifiers (Wilkins *et al.* 1994, 1996). Further, RBF ANNs can be trained relatively rapidly, and can discriminate unknown taxa upon which a network has not been trained. Haykin (1994) and Wilkins *et al.* (1994) provide detailed descriptions of RBF ANNs. Briefly, RBF networks have three layers of nodes or processing elements (Fig. 1). Flow cytometric data are input initially to the first layer, which serves merely

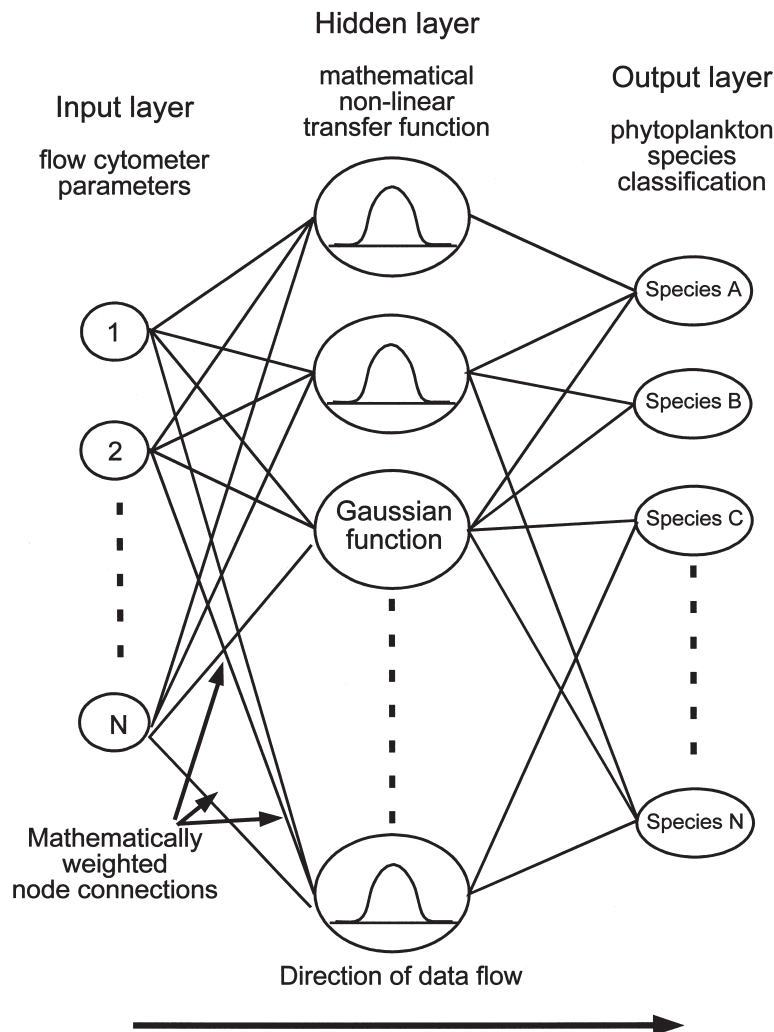


FIG. 1. – Architecture of a Radial Basis Function artificial neural network

to distribute these data to the hidden layer nodes (HLNs). The HLNs each represent a non-linear Gaussian basis function or kernel, the output value of which depends on the distance between the input data pattern and its centre. Several HLNs represent the data distributions for each species, and it is essential that there are sufficient HLNs to adequately fill the data input space. The outputs from the HLNs are passed to the output layer, which contains one node for each phytoplankton species. The output node with the highest value indicates the predicted identity of the input data.

ANNs learn from examples. To train an ANN, 400-500 patterns were drawn at random from the data files for each species and presented to the network together with the known identity of the pattern of flow cytometric characteristics. How well a network trains depends on the number of HLNs and a variety of other network parameters, which must be optimised by experimentally altering each of these factors in turn. The duration of training can be anything from a few minutes to half an hour or more, dependent on computer hardware, the number of species and complexity of the data. Once trained, the network must be tested using an independent data set drawn from the data files, and the identification performance evaluated by comparing the predicted identity of test patterns with their known identity. Misidentification results from overlap of character distributions, and can only be resolved by obtaining more and/or different characters. Within AIMS, data

analysis approaches have been initially developed to recognise target organisms using data of known origin, mono-specific phytoplankton cultures.

Phytoplankton species from the Plymouth Culture Collection (Marine Biological Association, UK) were chosen to provide wide taxonomic and size ranges (1-45 μm), of species found in European waters. Mono-specific cultures were grown at 15°C ($\pm 1^\circ\text{C}$) and were illuminated under constant light at 100-150 $\mu\text{mol m}^{-2} \text{s}^{-1}$. Batch cultures were grown in F/2 medium (Guillard and Ryther, 1962), in 250 ml polycarbonate flasks (Nalgene™) for several weeks before analysis and were sub-cultured every three to four days to maintain cultures in exponential growth.

Cultures were analysed by AFC using a Becton Dickinson FACSort™ flow cytometer equipped with a vertically polarised 15 mW argon ion laser emitting blue light at 488 nm. Data acquisition was triggered on chlorophyll fluorescence using laboratory cultures of *Micromonas pusilla* (1-3 μm) to set the lower analysis threshold. The flow cytometer detector array consisted of two fluorescence photo-multiplier tubes (PMTs), two light scatter PMTs and a photodiode for forward light scatter. Individual cell measurements were made for cellular forward light scatter, chlorophyll fluorescence ($>650 \text{ nm}$), phycoerythrin fluorescence (585 nm $\pm 21 \text{ nm}$), side scatter and depolarised light scatter (to enhance the discrimination of coccolithophores, Olson 1989). Samples were either run for three minutes at a flow rate of 107 $\mu\text{l min}^{-1}$ ($\pm 6 \mu\text{l}$), or until 10,000 events had been acquired. Analogue

TABLE 1. – Misidentification matrix for 20 species of marine phytoplankton. The numbers in bold running along the leading diagonal represent the percentage of events for a species being correctly identified. Off-diagonal numbers represent misidentifications of one species as another.

Species number and name	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1 <i>Aureodinium pigmentosum</i>	90	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	1	6	1	0
2 <i>Chaetoceros calcitrans</i>	0	99	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
3 <i>Chlorella salina</i>	1	1	81	0	0	3	1	2	7	0	0	0	0	0	0	0	0	4	0	0
4 <i>Chroomonas sp.</i>	0	0	0	97	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5 <i>Chroomonas salina</i>	0	0	0	2	92	0	0	0	0	0	1	1	3	1	0	0	0	0	0	0
6 <i>Chrysochromulina camella</i>	0	0	0	0	0	92	0	0	0	1	0	0	0	0	0	0	4	2	1	0
7 <i>Chrysochromulina chiton</i>	0	0	1	0	0	1	70	8	16	0	0	0	0	0	0	0	0	4	0	0
8 <i>Chrysochromulina cymbium</i>	1	0	4	0	0	0	13	65	14	0	0	0	0	0	0	0	0	3	0	0
9 <i>Chrysochromulina polylepis</i>	0	0	7	0	0	1	15	9	67	0	0	0	0	0	0	0	0	1	0	0
10 <i>Chlamydomonas reginae</i>	0	0	0	0	0	4	0	0	0	91	0	0	0	0	0	0	3	0	2	0
11 <i>Cryptomonas appendicularis</i>	0	0	0	0	0	0	0	0	0	0	98	0	2	0	0	0	0	0	0	0
12 <i>Cryptomonas calceiformis</i>	0	0	0	0	1	0	0	0	0	0	0	94	5	0	0	0	0	0	0	0
13 <i>Cryptomonas maculata</i>	0	0	0	0	0	0	0	0	0	0	1	5	94	0	0	0	0	0	0	0
14 <i>Cryptomonas reticulata</i>	0	0	0	0	3	0	0	0	0	0	0	0	0	97	0	0	0	0	0	0
15 <i>Cryptomonas rostellata</i>	0	0	0	0	0	0	0	0	0	0	0	0	1	99	0	0	0	0	0	0
16 <i>Dunaliella tertiolecta</i>	0	0	0	0	0	0	0	0	0	0	0	0	0	0	97	3	0	0	0	0
17 <i>Gymnodinium micrum</i>	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	87	3	6	3	0
18 <i>Gymnodinium simplex</i>	7	0	1	0	0	1	3	0	0	0	0	0	0	0	0	0	87	0	0	0
19 <i>Gymnodinium veneficum</i>	1	0	0	0	0	3	0	0	0	1	0	0	0	0	0	0	9	0	84	2
20 <i>Gyrodinium aureolum</i>	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	97

Overall performance (average of diagonal elements): 88.9% correct classification

TABLE 2. – Comparison of phytoplankton abundance as a proportion of total cell count by analytical flow cytometry (AFC) and a radial basis function artificial neural network (RBF) for three separate mixtures.

Species name	Mixture 1		Mixture 2		Mixture 3	
	AFC	RBF	AFC	RBF	AFC	RBF
<i>Amphidinium carterae</i>	0.51	0.49	0.08	0.06	0.00	0.00
<i>Cryptomonas maculata</i>	0.00	0.00	0.00	0.00	0.08	0.07
<i>Emiliana huxleyi</i>	0.00	0.01	0.00	0.01	0.12	0.13
<i>Hemiselmis rufescens</i>	0.11	0.10	0.00	0.00	0.08	0.08
<i>Hemiselmis virescens</i>	0.00	0.00	0.00	0.00	0.00	0.00
<i>Micromonas pusilla</i>	0.37	0.36	0.46	0.45	0.69	0.68
<i>Porphyridium purpureum</i>	0.00	0.01	0.41	0.40	0.00	0.00
<i>Prorocentrum minimum</i>	0.02	0.03	0.06	0.08	0.03	0.04

signals from the detectors were digitally converted and stored on computer as listmode data. Data files were then transferred to a PC by local area network or ethernet for analysis.

Networks were trained and were able to discriminate and quantify species in mixed assemblages. A typical example of some results is displayed in Table 1 as a misidentification matrix. Table 2 illustrates how the ANN-derived abundance estimates compare to abundance estimates obtained from classical flow cytometric analysis of dual parameter scatter plots for mixtures of 8 phytoplankton species. It can be seen that in this test, the RBF network estimates match the flow cytometric estimates very well.

Verifying the results of the networks is an area currently under development. One way to verify the identity of species predicted by an ANN is to transform the results into co-ordinates that can be used to produce sort decision boundaries on the flow cytometer. Cells can then be sorted from the sample using the flow cytometer's software and electronics and viewed by microscopy, with or without image analysis capabilities to enhance resolution. This approach was tested as part of an experimental workshop of the AIMS project. Simple mixtures containing five species of cultured phytoplankton were analysed on the Becton Dickinson™ FACSORT™ flow cytometer and identified by a trained RBF network. The clusters identified by the network as *Dunaliella tertiolecta* and *Tetraselmis suecica* were highlighted in scatter plots within the CytoWave flow cytometry acquisition and analysis software package and gates were drawn round the highlighted regions. These co-ordinates were then transferred to the FACSORT™ and were used as sort decision boundaries, one for each species. The sorted material was examined using microscopy and was identified as the two target species, thus proving that sorting can be performed in conjunction with artificial neural nets (see Fig. 2).

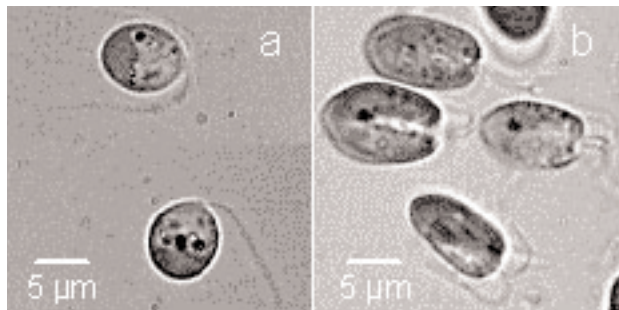


FIG. 2. – CCD camera images of *Dunaliella tertiolecta* and *Tetraselmis suecica* taken using the digital image analysis system of the University of Malaga, Spain.

There are several difficulties in extending the ANN methodology from laboratory mono-specific cultures in the laboratory to heterogeneous populations in the sea. Firstly, the number of classes in any natural seawater sample is unknown (i.e., the problem is unbounded). In addition, there may be taxa in the sample that have not been encountered by the network before. It is, therefore, essential to recognize these new taxa upon which it has not been trained and to label them as unknowns. Encountering unknowns is common to biology and biomedicine but not to other areas of technology, and has not been extensively investigated. RBF ANNs can deal with unknowns by applying constraints to outputs of HLN nodes or to output layer nodes (Morris and Boddy, 1996). To validate and determine the limitations constraint application within the AIMS project, mixed cultures and natural samples will be used. Once unknowns have been found in a sample they then need to be identified using the combination of flow cytometric sorting and microscopy described previously. Having made an identification the new taxa need to be added to the network. To add new species to a network for identification usually requires retraining of the network from scratch, which is time consuming and prone to error for

those not specialised in the use of ANNs. A possible solution is to provide a library of networks, each of which discriminates a single taxon from all others (Morris and Boddy, 1998). Numerous individual networks will be implemented during the project and the facility to train a net for a new taxon will be made available. Appropriate single species networks can then be combined to provide identifications. Identification of taxa may, however, be less important than identification and quantification of functional groups of organisms, and this may be relatively easily achieved by combining together appropriate groups of species.

Finally, obtaining truly representative training data is a major problem. This must cover the complete spectrum of biological variability within a species encountered in a natural environment. Data obtained from laboratory cultures do not accurately reflect characters of the same species in the field, because environmental conditions may dramatically affect cell characteristics measured by AFC. Ideally ANNs should be trained on data obtained from appropriate field samples. This will be achieved by identifying clusters of similar cells from AFC data (using unsupervised ANNs or statistical clustering methods), sorting these (as described above) and identifying the cells microscopically.

APPLICATION OF MOLECULAR PROBES OF TAXONOMIC AFFILIATION

Ribosomal RNA (rRNA) genes are ideal regions of the genome to use for the development of taxon-specific molecular probes. Not only do they contain regions of considerable sequence diversity (variable regions) but also regions that are evolutionarily conserved. These latter regions lend themselves to the design of universal primers for amplification of rRNA sequences using the polymerase chain reaction (PCR), as well as to the construction of family-, genus- and species-specific oligonucleotide probes. Although these attributes are also true for a number of evolutionarily conserved protein-encoding genes, the high copy number of rRNA molecules (up to 10^5 per cell) leads to vastly increased sensitivity and enables single cell analysis to be consistently achieved. Ribosomal RNA genes, especially from the small subunit (18S rRNA in eukaryotes, 16S rRNA in prokaryotes), are also commonly used in the reconstruction of evolutionary relationships among species: thus, a large number of sequences

are already available in public databases that can be used to develop taxon-specific probes.

Within the AIMS project, probes for the verification of phytoplankton taxa are being developed according to the following criteria: (a) relevance for European waters; (b) availability in culture collections for probe testing; (c) distribution of species over taxonomic groups and (d) different sizes and shapes of cells. These criteria make it possible not only to develop probes and general methodology required for the purpose of AIMS but also to compare their usefulness over a broad taxonomic base, with an emphasis on phytoplankton species that are important for European waters. The probes range from higher group level (e.g., Chlorophyll a+b = chlorophytes) down to the species level (e.g. *Heterocapsa triquetra*). Table 3 provides an overview of the algal and bacterial rRNA probes currently available or under development and used in this project (see also Peperzak *et al.*, 2000).

For designing these taxon specific rRNA probes for marine phytoplankton, a database of more than 330 aligned chlorophyll a+c algal 18S rRNA sequences was used. This alignment was analysed with the ARB program (Ludwig, TU München) to find regions of 18-20 nucleotides that were specific for different taxa, but also contained at least one mismatch to all other sequences in the database. The position and composition of the mismatch(es) were also taken into consideration as were GC content (percentage of guanine and cytosine bases) and Tm value of the probe sequence (temperature at which half of the DNA molecules are single stranded) in selecting an optimal probe (Stahl and Amann, 1991).

Potential probes were checked against all published sequences using RDP (Ribosomal Database Project, <http://www.cme.msu.edu/RDP/html>) (Maidak *et al.*, 1999) and BLAST (Basic Local Alignment Search Tool) search (<http://www.ncbi.nlm.nih.gov/BLAST>) (Altschul *et al.*, 1990) to find possible matches to sequences not included in our database. Furthermore, the probes were analysed by checking for their ability to form internal loops and self-hybridisation, which may prevent them from binding to their target molecule. Finally, the probe was localised on the two-dimensional structure of the rRNA molecule. It is possible that not all regions of the rRNA are accessible to the probes because of the molecule's secondary structure or because some parts of the molecule are covered by ribosomal proteins. Fuchs *et al.* (1998) found that in *E. coli*, probes in some regions of the

TABLE 3. – Algal and bacterial probes used in the AIMS project. The table lists not only probes developed within the framework of the AIMS project, but also some from other working groups.

Higher Group Probes	Eukaryotes (Lim <i>et al.</i> , 1993) Eubacteria (Amann <i>et al.</i> , 1990) Chlorophyta (Simon <i>et al.</i> , 1995, 2000) Non-Chlorophyta (Simon <i>et al.</i> , 1995)
Class Level Probes	Alpha-Proteobacteria (Manz <i>et al.</i> , 1992) Beta-Proteobacteria (Manz <i>et al.</i> , 1992) Gamma-Proteobacteria (Manz <i>et al.</i> , 1992) Dinophyceae (Medlin and Groben, unpubl.) Pelagophyceae (Simon <i>et al.</i> , 2000) Prymnesiophyceae (Lange <i>et al.</i> , 1996; Simon <i>et al.</i> , 2000)
Order Level Probes	Pennate Diatoms (Medlin, unpubl.)
Clade Level Probes	(Probes specific for phylogenetic clades above or inside genera) Toxic Chrysochromulina and Prymnesium clade (Simon <i>et al.</i> , 1997) Non-toxic Chrysochromulina clade (Simon <i>et al.</i> , 1997) Cryptophyta, probes for six different clades (Groben and Medlin, unpubl.) Pelagophyceae clade (Simon <i>et al.</i> , 2000) Prorocentrum, probes for two different clades (Lange and Medlin, unpubl.)
Genus Level Probes	<i>Chaetoceros</i> (Groben and Medlin, unpubl.) <i>Phaeocystis</i> (Lange <i>et al.</i> , 1996)
Species Level Probes	<i>Alexandrium tamarense</i> (Miller and Scholin, 1998; additional unpubl. probes) <i>Chrysochromulina polylepis</i> (Simon <i>et al.</i> , 1997) <i>Emiliania huxleyi</i> (Moon-van der Staay <i>et al.</i> 2000, Groben and Medlin, unpubl.) <i>Gymnodinium mikimotoi</i> species complex (Groben, Vrieling, Olsen, Medlin, unpubl.; Sako <i>et al.</i> , 1998) <i>Heterocapsa triquetra</i> (Groben and Medlin, unpubl.) <i>Phaeocystis globosa</i> (Groben and Medlin, unpubl.) <i>Prorocentrum lima</i> (Lange, Groben and Medlin, unpubl.) <i>Prorocentrum micans</i> (Lange and Medlin, unpubl.) <i>Prorocentrum minimum</i> (Lange, Groben and Medlin, unpubl.) <i>Skeletonema costatum</i> (Groben and Medlin, unpubl.)

rRNA sequence gave much weaker hybridisation signals than others and presumably the sites conferring low signal were inaccessible because they were blocked by ribosomal proteins. Even if these results are not directly applicable to eukaryotic (algal) species, it appears that some regions are preferred for designing probes. Any probes passing these different tests were synthesised and used in hybridisation experiments to validate probe specificity.

Routine testing, for probe specificity, involves a 2-tiered approach. First, probes were tested against a broad taxonomic base using a chemiluminescent detection system in which probes were hybridised with filter-bound, PCR-amplified rDNA from different algal species (“dot blots”). For these tests, DNA from target species (algae for which the probe was designed) were blotted and fixed onto positive-charged nylon membrane. DNA from non-target

species was similarly used as a negative control. Non-target species were either taxonomically closely related to the target organisms or showed similarities to the target in the probe-binding region of the rRNA molecule. A specific probe should only give signals with DNA from its target species, but not with those from non-target species. Probes that showed only specific binding to their target species in the dot blot hybridisation were used for *in situ* hybridisation experiments. The *in situ* hybridisation was designed to address the problems encountered as the probe penetrates the cell and hybridises with the native configuration of the rRNA molecule.

For hybridisation of whole algal cells, the oligonucleotide probes were labelled with the fluorochromes fluorescein isothiocyanate (FITC) or Cy3. The same species were used for *in situ* hybridisation tests of probe specificity as in the DNA dot blot experiments. For this, ethanol-treated cells are hybridised with the probes because chlorophyll is soluble in ethanol. Thus the autofluorescence of the cells was reduced, preventing it from masking the probe signal. Cells were incubated for 1-2 hours with the probes in a hybridisation buffer at a specific, probe-dependent temperature. Afterwards, the cells were washed three times in the same buffer at a slightly higher temperature to remove binding of the probe to non-specific targets.

For fluorescence microscopy, the samples were resuspended in Citifluor (Citifluor Ltd., Canterbury, UK) to prevent fading of probe signals and counterstained with DAPI for better visualisation of non-labelled cells. Specificity testing with fluorescence microscopy includes hybridisation to target and non-target algae at different hybridisation temperatures and different ratios of labelled to non-labelled probes as competitors, to show the specificity of probe binding. Different hybridisation temperatures were used because temperature dependent, specific binding of probes has been demonstrated as signal strength decreases with increasing temperature. The use of fluorescently-labelled rRNA probes in combination with fluorescence microscopy or with flow cytometry can result in specific detection of a single species or a higher taxon (Simon *et al.*, 1995, 1997, 2000; Lange *et al.*, 1996; Miller and Scholin, 1998).

Approximately one third of all probes used in the AIMS project have been successfully tested with *in situ* hybridisation and flow cytometry (e.g., Simon *et al.*, 1995, 2000; Lange *et al.*, 1996). These are mainly the ones that are specific for classes and higher groups of phytoplankton. The remaining probes are

species-specific probes, whose specificity has only been tested to date in DNA dot blot analyses.

The development and application of specific rRNA probes offers great potential for the analysis of phytoplankton. As with all new methods, a number of questions remain, such as how these probes will perform in field tests. These questions will be addressed within the AIMS project, using experiments in mesocosms and research cruises. Also, new species must be tested with existing probes to confirm that probe specificity is maintained. The use of rRNA probes is a valuable tool for the identification and characterisation of phytoplankton populations either in combination with flow cytometry or with fluorescence microscopy. With this tool it is, for example, possible to monitor harmful algal blooms (Brenner *et al.*, in prep.) and to characterise species that cannot be cultured, e.g., the dinoflagellate *Dinophysis*. Molecular probes will also be able to characterise taxa that are difficult to analyse by flow cytometry, such as colonial and chain-forming species, as well as separating species with similar optical properties.

INTEGRATING IDENTIFICATION AND CHARACTERISATION OF CELLS: THE AIMS SOFTWARE AND DATABASE

The main end product of the AIMS project will be an integrated software package combining flow cytometry acquisition and visualisation software (CytoWave), artificial neural network software (AIMSNet), and algorithms to convert raw flow cytometric data into inherent properties of cells, such as, cell size, volume, pigment content and refractive index (CellStat).

A database of simultaneous measurements of biomass, pigment content, light absorption, light scattering and cell size has been compiled from measurements on laboratory cultures and mesocosm experiments and will also include data from a sea-going research cruise in 2000. This database, called AIMSBase, will also contain flow cytometric analyses and is accessible via internet (www.flowcytometry.org). AIMSBase allows for the comparison of flow cytometric measurements of forward scattering, right angle scattering, time of flight and fluorescence with chemical and optical measurements in order to develop algorithms for estimating pigment and carbon contents of microbial cells from flow cytometric measurements.

A number of different analytical flow cytometers is being used within this project because differences in optical geometry can lead to wide variations in scattering signals and because the final software will be designed for use with multiple flow cytometer configurations. Measurements have been made using general purpose (Becton Dickinson™ FACScan™ and FACSort™, Coulter XL™) and specialised (Cyto-Buoy™, Dubelaar *et al.*, 1999) flow cytometers.

The data acquisition and visualisation software includes integrated artificial neural networks for data analysis. This integration should permit near real-time ANN analysis of flow cytometric data. The results of an ANN analysis can be displayed graphically by means of standard representations of the data such as dot plots and histograms, as well as in tabular form, listing taxa present with an estimate of their abundance.

The software allows trained ANNs to be generated for different combinations of species or subpopulations. The use of combinations of networks is also foreseen, allowing the rapid construction of networks to accommodate novel combinations of species by combining ANNs drawn from a library of existing pre-trained networks, each specialised to a particular group of species or set of environmental conditions.

The progress in the application of ANNs, combined with the discrimination power of rRNA probes and the estimation of inherent properties of cells will greatly improve the application of flow cytometry for the automated analysis of phytoplankton in field samples and the use of flow cytometry for monitoring purposes.

ACKNOWLEDGEMENTS

The authors wish to thank all partners of the AIMS project for their help and co-operation. This work is supported by the EU grant MAS3-CT97-0080. This is AWI contribution no. 1741 and forms part of the work of the Core Strategic Research Project DYME of the CCMS Plymouth Marine Laboratory.

REFERENCES

- Altschul, S.F., W. Gish, W. Miller, E.W. Myers and D.J. Lipman. – 1990. Basic local alignment search tool. *J. Mol. Biol.*, 215: 403-410.
- Amann, R.L., B.J. Binder, R.J. Olson, R.J. Chisholm, R. Devereux and D.A. Stahl. – 1990. Combination of 16S rRNA-targeted oligonucleotide probes with flow cytometry for analyzing

- mixed microbial populations. *Appl. Environ. Microbiol.*, 56: 1919-1925.
- Amann, R.L., W. Ludwig and K.-H. Schleifer. – 1995. Phylogenetic identification and *in situ* detection of individual cells without cultivation. *Microbiol. Rev.*, 59: 143-169.
- Barer, R. and S. Joseph. – 1954. Refractometry of living cells. Part 1. Basic principles. *Q. J. Microscopical Sci.*, 95: 399-423.
- Barlow, R.G., R.F.C. Mantoura and D.G. Cummings. – 1999. Monsoonal influence on the distribution of phytoplankton pigments in the Arabian Sea. *Deep-Sea Res. II*, 46: 677-699.
- Barlow, R. G., R.F.C. Mantoura, M.A. Gough and T.W. Fileman. – 1993. Pigment signatures of the phytoplankton composition in the northern Atlantic during the 1990 spring bloom. *Deep-Sea Res.*, 40: 459-477.
- Bidigare, R. R. and M.E. Ondrusek. – 1996. Spatial and temporal variability of phytoplankton pigment distributions in the central equatorial Pacific Ocean. *Deep-Sea Res. II*, 43: 809-833.
- Boddy, L., C.W. Morris, M.F. Wilkins, G.A. Tarran and P.H. Burkill. – 1994. Neural network analysis of flow cytometric data for 40 marine phytoplankton species. *Cytometry*, 15: 283-293.
- Brenner, J. *et al.* – Manuscript in preparation. Quantification of algal blooms (*Alexandrium tamarense*) with *in situ* hybridization, flow cytometry and neural networks.
- Burkill, P.H. and R.F.C. Mantoura. – 1990. The rapid analysis of single marine cells by flow cytometry. *Phil. Transact. Royal Soc. Ser. A*, 333: 99-112.
- Carr, M.R., G.A. Tarran and P.H. Burkill. – 1996. Discrimination of marine phytoplankton species through the statistical analysis of their flow cytometric signatures. *J. Plankton Res.*, 18: 1225-1238.
- Culverhouse, P.F., R.G. Simpson, R. Ellis, J.A. Lindley, R. Williams, T. Parisini, B. Reguera, I. Bravo, R. Zoppoli, G. Earnshaw, H. McHall and H. Smith. – 1996. Automatic identification of field-collected dinoflagellates by artificial neural network. *Mar. Ecol. Prog. Ser.*, 139: 281-287.
- Culverhouse, P.F., R. Ellis, R.G. Simpson, R. Williams, R.W. Pierce and J.T. Turner. – 1994. Automatic categorisation of five species of Cymatocylis (Protozoa, Tintinnida) by artificial neural network. *Mar. Ecol. Prog. Ser.*, 107: 273-280.
- Demers, S., J. Kim, P. Legendre and L. Legendre – 1992. Analysing multivariate flow cytometric data in aquatic sciences. *Cytometry*, 13: 291-298.
- Dubelaar, G.B.J., P.L. Gerritzen, A.E.R. Beeker, R.R. Jonker, K. Tangen – 1999. Design and first results of CytoBuoy: A wireless flow cytometer for *in situ* analysis of marine and fresh waters. *Cytometry*, 37: 247-254.
- Frankel, D.S., S.L. Frankel, B.J. Binder, R.F. Vogt – 1996. Application of neural networks to flow cytometry data analysis and real-time cell classification. *Cytometry*, 23: 290-302.
- Frankel, D.S., R.J. Olson, S.L. Frankel and S.W. Chisholm. – 1989. Use of a neural network computing system for analysis of flow cytometric data of phytoplankton populations. *Cytometry*, 10: 540-550.
- Fuchs, B.M., G. Wallner, W. Beisker, I. Schwippl, W. Ludwig and R. Amann. – 1998. Flow cytometric analysis of the *in situ* accessibility of *Escherichia coli* 16S rRNA for fluorescently labeled oligonucleotide probes. *Appl. Environ. Microbiol.*, 64: 4973-4982.
- Giovanoni, S.J., T.D. Mullins and K.G. Field -1995. Microbial diversity in oceanic systems: rRNA approaches to the study of unculturable microbes. In: I.R. Joint (ed.) *Molecular Ecology of Aquatic Microbes*. NATO ASI Series G 38, pp. 217-248. Springer-Verlag, Berlin.
- Guillard, R.R.L and J.H. Ryther. – 1962. Studies on marine planktonic diatoms I. *Cyclotella nana* (Hustedt) and *Detonula confervacea* (Cleve) Gran. Canadian. *J. Microbiol.*, 8: 229-239.
- Haykin, S. – 1994. Neural Networks: a comprehensive foundation. Maxwell MacMillan International, New York.
- Hofstraat, J.W., W.J.M. van Zeijl, M.E.J. de Vreeze, J.C.H. Peeters, L. Peperzak, F. Colijn and T.W.M. Rademaker – 1994. Phytoplankton monitoring by flow cytometry. *J. Plankton Res.*, 16: 1197-1224.
- Jonker, R.R., J.T. Meulemans, G.B.J. Dubelaar, M.F. Wilkins and J. Ringelberg. – 1995. Flow cytometry: a powerful tool in analysis of biomass distributions of phytoplankton. In: L.R. Mur *et al.* (eds.), *IAWQ SIL Conference on Selection Mechanisms controlling biomass distributions*, 32, pp. 177-182.
- Lange, M., L. Guillou, D. Vaultot, N. Simon, R.I. Amann, W. Ludwig and L.K. Medlin. – 1996. Identification of the class Prymnesiophyceae and the genus *Phaeocystis* with ribosomal RNA-targeted nucleic acid probes detected by flow cytometry. *J. Phycol.*, 32: 858-868.
- Li, W.K.W., T. Zohary, Y.Z. Yacobi and A.M. Wood. – 1993. Ultraphytoplankton in the eastern Mediterranean Sea: towards deriving phytoplankton biomass from flow cytometric measurements of abundance, fluorescence and light scatter. *Mar. Ecol. Prog. Ser.*, 102: 79-87.
- Lim, E.L., L.A. Amaral, D.A. Caron and E.F. DeLong. – 1993. Application of rRNA-based probes for observing nanoplanktonic protists. *Appl. Environ. Microbiol.*, 59: 1647-1655.
- Maidak B.L., J.R. Cole, C.T. Parker, jr., G.M. Garrity, N. Larsen, B. Li, T.G. Lilburn, M.J. McCaughey, G.J. Olsen, R. Overbeek, S. Pramanik, T.M. Schmidt, J.M. Tiedje and C.R. Woese. – 1999. A new version of the RDP (Ribosomal Database Project). *Nucl. Acids Res.*, 27: 171-173.
- Manz, W., R. Amann, W. Ludwig, M. Wagner and K.-H. Schleifer. – 1992. Phylogenetic oligodeoxynucleotide probes for the major subclasses of proteobacteria: problems and solutions. *Syst. Appl. Microbiol.*, 15: 593-600.
- Miller, P.E. and C.A. Scholin. 1998. Identification and enumeration of cultures and wild *Pseudo-Nitzschia* (Bacillariophyceae) using species-specific LSU rRNA-targeted fluorescent probes and filter-based whole cell hybridization. *J. Phycol.*, 34: 371-382.
- Moon-van der Staay, S. Y., G.W.M van der Staay, L. Guillou, H. Claustre, L.K. Medlin and D. Vaultot – 2000. Abundance and diversity of Prymnesiophyceae in the picoplankton community from the equatorial Pacific Ocean inferred from 18S rDNA sequences. *Limnol. Oceanogr.*, in press.
- Morel, A. – 1991. Optics of marine particles and marine optics. In: S. Demers (ed.), *Particle analysis in Oceanography NATO ASI Series*, G 27, pp 141-188. Springer Verlag, Berlin, New York.
- Morris, C.W. and L. Boddy. – 1996. Classification as unknown by RBF networks: discriminating phytoplankton taxa from flow cytometry data. In: C.H. Dagli *et al.* (eds.), *Intelligent Engineering Systems through Artificial Neural Networks*, 6. ASME Press.
- Morris, C.W. and L. Boddy. – 1998. Partitioned RBF networks for identification of biological taxa: discrimination of phytoplankton from flow cytometry data. In: C.H. Dagli *et al.* (eds.), *Intelligent Engineering Systems through Artificial Neural Networks*, 8. ASME Press.
- Olson, R.J., E.R. Zettler *et al.* – 1989. Discrimination of eukaryotic phytoplankton cell types from light scatter and autofluorescence properties measured by flow cytometry. *Cytometry*, 10: 636-643.
- Peperzak, L, E.G. Vrieling, B. Sandee and T. Rutten. – 2000. Immuno flow cytometry in marine phytoplankton research. *Sci. Mar.*, 64(2): 165-181.
- Rice, J., M.A. Sleight, P.H. Burkill, G.A. Tarran, C.D. O'Connor and M.V. Zubkov. – 1997a. Flow cytometric analysis of characteristics of hybridization of species-specific fluorescent oligonucleotide probes to rRNA of marine nanoflagellates. *Appl. Environ. Microbiol.* 63: 938-944.
- Rice, J., C.D. O'Connor, M.A. Sleight, P.H. Burkill, I.G. Giles, and M.V. Zubkov. – 1997b. Fluorescent oligonucleotide rDNA probes that specifically bind to a common nanoflagellate, *Paraphysomonas vestita*. *Microbiol.*, 143: 1717-1727.
- Sako, Y., M. Rokushima, M. Yamaguchi, Y. Ishida and A. Uchida. – 1998. Phylogenetic analysis and molecular identification of red tide dinoflagellate *Gymnodinium mikimotoi* using ribosomal RNA genes. In: B. Reguera, J. Blanco, M.L. Fernandez and T. Wyatt (eds.): *Harmful Algae*, pp. 295-298. Xunta de Galicia and Intergovernmental Oceanographic Commission of UNESCO.
- Simon, N., J. Brenner, B. Edvardsen and L. Medlin. – 1997. The identification of Chrysochromulina and Prymnesium species (Haptophyta, Prymnesiophyceae) using fluorescent or chemiluminescent oligonucleotide probes: a means for improving studies on toxic algae. *Eur. J. Phycol.*, 32: 393-401.
- Simon, N., L. Campbell, E. Ornlöf, R. Groben, L. Guillou, M. Lange and L.K. Medlin. – 2000. Oligonucleotide probes for the identification of three algal groups by dot blot and fluorescent whole-cell hybridization. *J. Euk. Microbiol.*, 47: 76-84.
- Simon, N., N. LeBot, D. Marie, F. Partensky and D. Vaultot. – 1995. Fluorescent *in situ* hybridisation with rRNA-targeted oligonucleotide probes to identify small phytoplankton by flow cytometry. *Appl. Environ. Microbiol.*, 61: 2506-2513.
- Spinrad, R.W. and J.F. Brown. – 1986. Relative real refractive

- index of marine microorganisms: a technique for flow cytometric estimation. *Appl. Optics*, 25: 1930-1935.
- Stahl, D.A. and R. Amann. – 1991. Development and application of nucleic acid probes. In: E. Stackebrandt and M. Goodfellow (eds.), *Nucleic Acid Techniques in Bacterial Systematics*, pp. 205-248. John Wiley & Sons Ltd., Chichester, New York, Brisbane, Toronto, Singapore.
- Stramski, D., A. Morel and A. Bricaud. – 1988. Modeling the light attenuation and scattering by spherical phytoplankton cells: retrieval of the bulk refractive index. *Applied Optics*, 27: 3954-3956
- Stramski, D. and R.A. Reynolds. – 1993. Diel variations in the optical properties of a marine diatom. *Limnol. Oceanogr.*, 38: 1347-1364
- Veldhuis, M.J.W., T.L. Cucci and M.E. Sieracki. – 1997. Cellular DNA content of marine phytoplankton using two new fluorochromes: taxonomic and ecological implications. *J. Phycol.*, 33: 527-541.
- Wilkins, M.F., C.W. Morris and L. Boddy. – 1994. A comparison of radial basis function and backpropagation neural networks for identification of marine phytoplankton from multivariate flow cytometry data. *CABIOS*, 10: 285-294.
- Wilkins, M.F., L. Boddy, C.W. Morris and R. Jonker. – 1996. A comparison of some neural and non-neural methods for identification of phytoplankton from flow cytometry data. *CABIOS*, 12: 9-18.
- Williams, R., H. McHall, R.W. Pierce and J.T. Turner. – 1994. Speciation of the tintinnid genus *Cymatocylis* by morphometric analysis of the loricae. *Mar. Ecol. Prog. Ser.*, 107: 263-272.