Appendices

In order to support the readers of this book in their endeavor to understand and utilize neural networks, we provide them with additional material in electronic form. Rather than including this material on diskette or CD-ROM in this book we have decided to install a website that contains this additional information. This allows us to constantly expand and update this material.

The URL for this website is

http://www2.ccc.uni-erlangen.de/ANN-book/

The additional information includes access to programs, data sets, presentation materials, tutorials, and publications.

A Programs

A.1 A Hopfield Program, HOPF

A demo-program for the application of a Hopfield networks is provided on the website. The example detailed in Section 4.4 can be worked through explicitly. The program HOPF.EXE and a data file containing the patterns shown in Figures 4-2 and 4-5 can be downloaded and run on a personal computer. Furthermore, the user can input his/her own images in the format of 3x3, 4x4, 5x5, 6x6, 7x7, or 8x8 pixels. The program can handle up to 64 images.

A.2 An Adaptive Bidirectional Associative Memory Program, ABAM

A simulator program for an ABAM network is provided. The user can study the example explained in Section 5.5 by himself. The program ABAM.EXE and the two data files used to produce the results given in Figure 5-9 and Figure 5-10, respectively, can be downloaded from the website. The program can also be used to study data of one's own. Up to 20 different pairs of input images / output targets can be handled.

A.3 A Kohonen Network Simulator

This is a Kohonen network program that has been augmented with various software tools for chemical applications:

- a browser for chemical structures *csbr*
- a tool for the visualization of Kohonen maps

It is therefore particularly helpful for the study of chemical data sets.

A.4 A Suite of Neural Network Programs, SNNS

A program package that includes a large variety of neural network models has been developed by the group of Prof. Andreas Zell, first at the University of Stuttgart and now at the University of Tübingen, Germany.

It can be accessed and downloaded from the website at:
http://www.informatik.uni-stuttgart.de/ipvr/bv/projekte/snns/snns.html

A.5 Telecooperation in Spectroscopy, TeleSpec

The methods presented in Sections 18.7 - 18.10 for the simulation of infrared spectra and the prediction of the 3D structure of a molecule from its infrared spectrum have been implemented on the internet on the TeleSpec site at *http://www2.ccc.uni-erlangen.de/research/ir/*.

For the simulation of an infrared spectrum, the user only has to input the constitution of a molecule. For this purpose, a molecule editor is provided. A 3D model of the molecule will automatically be generated by CORINA and transformed either into a 3D-MoRSE code or an RDF code. A counter-propagation network will be trained with similar structures and their corresponding infrared spectra, as retrieved from the SpecInfo© database. Input of the query structure into this counter-propagation network will produce a simulated infrared spectrum.

By the same token, the reverse mode of utilizing a counter-propagation network has been implemented. On input of an infrared spectrum, a 3D structure for the corresponding molecule is proposed. The various steps in these methods are briefly explained on the website; an interactive tutorial is provided.

In addition, access to a publicly available database of infrared spectra is provided. The database has been compiled by the users of the TeleSpec project and everybody is invited to send in her/his infrared spectra and associated structures to contribute to this service to the scientific community that is to the benefits of anybody interested in using infrared spectra for structure identification and elucidation.

A.6 The 3D Structure Generator CORINA

Molecules are three-dimensional objects and, therefore, many physical, chemical, and biological properties depend on the 3D structure. This has been demonstrated in various places throughout this book. Several methods for the encoding of the 3D structure of molecules have been presented.

In order to allow the investigation of the relationships between a property and the 3D structure, access to the Cartesian coordinates of the atoms of a molecule is necessary. Such an access is provided by the universal automatic 3D structure generator CORINA. Access to the 3D structure of up to 1,000 compounds is provided for free at *http://www2.ccc.uni-erlangen.de/software/corina/free_struct.html.* For the conversion of larger data set of molecules the company Molecular Networks GmbH should be contacted at *http://www.mol-net.de/.*

A.7 Calculation of Physicochemical Descriptors, the PETRA System

A variety of empirical methods for the calculation of all-important physicochemical effects, such as charge distribution, inductive, resonance, or polarizability effect has been developed since the mid-seventies in the research group of Prof. J. Gasteiger. These methods have been collected in the program package PETRA (Parameter Estimation for the Treatment of Reactivity Applications). The benefits of the results from these methods have been demonstrated at various places in this book. Details on these methods and access to the manual and the program is provided on the website at: *http://www2.ccc.uni-erlangen.de/software/petra/.*

B Data Sets

B.1 Italian Olive Oils

The data on the 512 samples of Italian olive oils, studied in Chapter 10 are stored on the website and can be downloaded. The data comprise the analysis of the contents of eight fatty acids in olive oil samples and give the region of origin.

We thank Prof. M. Forina, University of Genova, Italy for providing us with this data set and giving us permission to further distribute these data.

B.2 Steroids Binding to the CBG Receptor

A data set of 31 steroids, binding to the corticosteroid binding globulin (CBG) receptor was studied in Sections 13.5 to 13.8 and 19.5.

This data set is widely used; c.f. E. A. Coats "The CoMFA Steroids as a Benchmark Data Set for Development of 3D QSAR Methods" in *3D QSAR in Drug Design,* Vol. 3, Eds.: H. Kubinyi, G. Folkers and Y. C. Martin, Kluwer/ESCOM, Dordrecht, NL, 1998, pp. 199 – 214.

However, the structural formulas in printed publications and in electronic data file all contained errors. After going through the original publications, all coding errors have been eliminated (M. Wagener, J. Sadowski and J. Gasteiger, *J. Am. Chem. Soc.* **117** (1995) 7769 – 7775). The data set is here made accessible with the structures coded as connection tables with stereochemical information in the Molfile format. Furthermore, the activity values and the activity classification for binding to the CBG receptor are given (cf. Table 13.2).

B.3 Combinatorial Library

In order to also provide a larger data set to the community, the combinatorial libraries studied in Sections 20.5 – 20.6 comprising 65,341 dimethylxanthene, 11,191 cubane, and 11,191 adamantane derivatives are made accessible. Because of the size of the data set we refrain from giving the connection tables as Molfile (ASCII) data sets.

Rather, we provide for each of the compounds the 12 autocorrelation coefficients obtained from the molecular electrostatic potential on the van der Waals surface.

C Presentation Material

We have prepared two slide shows, one on the basic concepts of neural networks, and another one on various types of applications. These slide shows have been prepared quite some time ago with the program DrHalo. This program and the data sets of the slide shows can be downloaded and run on any PC under DOS or Windows.

D Publications

A list of publications on the application of neural networks to chemical problems from our two research groups is provided. Furthermore, for some of these publications in printed form the copyright has been raised so that these papers can be reproduced in full form.

E Tutorials

A tutorial on the use of counter-propagation networks is contained on the website.