ELSEVIER

Neural networks as a tool for compact representation of *ab initio* molecular potential energy surfaces

Erwin Tafeit, Willibald Estelberger, Renate Horejsi, Reinhard Moeller, Karl Oettl, Karoline Vrecko, and Gilbert Reibnegger

Medizinisch-Chemisches Institut und Pregl-Laboratorium, Karl-Franzens-Universität Graz, Graz, Austria

Ab initio *quantum chemical calculations of molecular properties such as, e.g., torsional potential energies, require massive computational effort even for moderately sized molecules, if basis sets with a reasonable quality are employed. Using* ab initio *data on conformational properties of the cofactor (6*R*,1′*R*,2′*S*)-5,6,7,8-tetrahydrobiopterin, we demonstrate that error backpropagation networks can be established that efficiently approximate complicated functional relationships such as torsional potential energy surfaces of a flexible molecule. Our pilot simulations suggest that properly trained neural networks might provide an extremely compact storage medium for quantum chemically obtained information. Moreover, they are outstandingly comfortable tools when it comes to making use of the stored information. One possible application is demonstrated, namely, computation of relaxed torsional energy surfaces.*

Keywords: supervised learning, molecular conformation, torsional energy

INTRODUCTION

The search for accurate electronic wave functions of polyatomic molecules uses mainly the molecular orbital method. These wave functions are dependent in a complicated manner on bond distances, bond angles, and dihedral (torsional) angles of rotations about single bonds. Wave functions are generally not available in analytical form (many-body problem). Even for moderately sized molecules, quantum chemical calculations of molecular properties on the *ab initio* level of quality are extremely time consuming if basis sets with reasonable quality and flexibility are employed. When biologically relevant molecules of tractable size are investigated by such theoretical methods, the computational effort often forces one to content oneself with an estimation of the optimum geometry and the corresponding energy of the molecule. Knowing only the optimum geometry of the molecule, however, is in general not satisfactory to judge its behavior in situations of practical interest; at least rough knowledge of the complicated dependence of the potential energy (or other molecular properties) on variations of the structural variables of the molecules is frequently desirable. For example, for proper binding of a molecule to a receptor it might be necessary that the molecule undergoes slight geometric rearrangement, for example by rotations about single bonds. Therefore, information is desirable about geometrical deviations from optimum geometry that are energetically accessible at a given temperature.

Artificial neural networks are a group of computational approaches to master complicated tasks in the absence of analytical solutions to a given problem. Also termed "connectionist systems," they are inspired by and loosely resemble real neural systems: in particular, they process information in a parallel manner by use of massively interconnected individual computational elements. Basic concepts of such computing schemes date back to the 1940s.[1] After an initial burst of enthusiasm,[2] seemingly intractable problems occurred and only a few pioneers continued work in further developments.[3] The demonstration that principles of theoretical physics of many-particle systems are applicable to certain network topologies[4] and the invention of efficient learning schemes[5] during the 1980s led to a remarkable renaissance in the field. Meanwhile, neural networks are used in different fields of applications, and there are a huge number of network architectures and algorithms available.

Error back propagation networks[5] are considered to represent a flexible and generally applicable family of neural networks, having been exploited for classification problems as well as for the representation of functional relationships. We have investigated the suitability of neural networks of this type to represent, in a compact manner, the torsional

Address reprint requests to Dr. Reibnegger at the Medizinisch-Chemisches Institut und Pregl-Laboratorium, Karl-Franzens-Universität Graz, Harrachgrasse 21/II, A-8010 Graz, Austraia.
Received 8 November 1995; accepted 21 November 1995.

655 Avenue of the Americas, New York, NY 10010

0263-7855/96/$15.00
SSDI 0263-7855(95)00087-9

potential energy of a biologically important molecule. The networks were adapted to their purpose by training data stemming from *ab initio* quantum chemical computations of this molecule. The aims of our study were to investigate the feasibility of such an approach in principle, to determine its limitations, and to define a strategy for treating realistic problems.

METHODS

Data

We have previously reported results of an extensive conformational investigation of tetrahydrobiopterin.[6] This pteridine derivative is an essential cofactor of phenylalanine 4-, tyrosine 3-, and tryptophan 5-monooxygenases.[7,8] In addition, glyceryl ether cleavage[9] and nitric oxide formation[10,11] by conversion of arginine to citrulline require tetrahydrobiopterin. The structure of the biologically active stereoisomer, (6*R*,1′*R*,2′*S*)-5,6,7,8-tetrahydrobiopterin, is shown in Figure 1a. Also shown (Figure 1b) is the most stable conformer of the molecule, in which the 1′,2′-dihydroxypropyl side chain is in the axial conformation with respect to the tetrahydropyrazine moiety. The molecule possesses two interesting flexible bonds, namely, between C-6 of the ring system and C-1 (torsional angle Φ), and between C-1′ and C-2′ of the side chain (torsional angle Ψ). To investigate the torsional potential energy with respect to these flexible bonds, the atoms comprising the ring system were held fixed, and the torsional angles of the two flexible bonds were varied systematically. Two rounds of angle variation were performed: first, starting at Φ = Ψ = 0°, the angles were varied in 45° steps, and second, starting at Φ = Ψ = 22.5°, the angles were again varied in 45° steps. At each position of the two torsional angles, the energy of the molecule was computed by an *ab initio* quantum chemical technique: the closed-shell Hartree–Fock Hamiltonian was used on the self-consistent field level, employing the 3-21G basis set, which yields reasonable results for pteridines.[12] For these computations, the program Gaussian (Gaussian, Pittsburgh, PA) was employed. The results of the energy computations in dependence on Φ and Ψ are also included in Figure 1c as a contour plot.

This potential energy surface suffers from an important deficiency because the molecule at any specific constellation of Φ and Ψ was not allowed to relax, simply because of the huge computational efforts: tetrahydrobiopterin is a relatively large molecular system in terms of an *ab initio* calculation of the quality level employed, and one single-point calculation of the molecule at any distinct geometry consumed approximately 4.5 hr of CPU time on an IRIS Indigo workstation (Silicon Graphics, Mountain View, CA). Calculation of a relaxed torsional potential energy surface would have required geometry optimization at each constellation of Φ and Ψ, a task vastly too large with respect to computer costs (each geometry optimization might require up to 100 or even more single-point calculations). The contour plot in Figure 1c thus represents upper bounds of the rotational energy barriers; in fact, the molecule is expected to be more flexible than indicated.

General network design and architecture

Feed-forward neural networks were used to approximate the quantum chemically established torsional potential energy surface of tetrahydrobiopterin depicted in Figure 1c. The input layer of all networks consisted of two computational units (termed "neurons" for simplicity henceforth). These neurons served as receivers of the inputs into the networks, i.e., the torsional angles Φ and Ψ. The output layer consisted of only one neuron, and its output was identified as

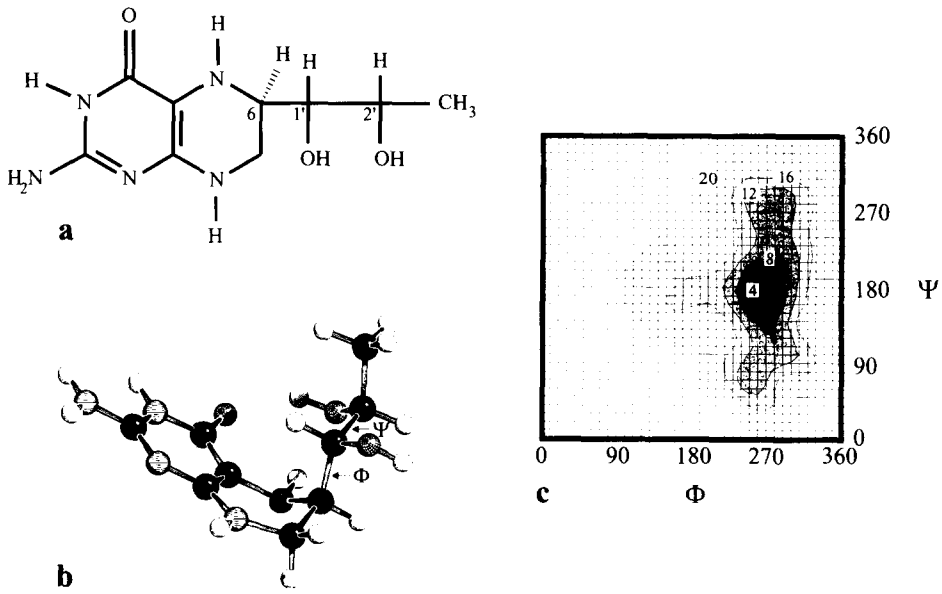

Figure 1. Chemical structure of (6R,1′R,2′S)*-5,6,7,8-tetrahydrobiopterin. (a) Configuration of the biologically active stereoisomer; (b) conformation with minimal energy; (c) torsional potential energy surface,* ab initio *computation (the numbers in the contour plot denote torsional energies, in kilocalories per mole). The torsional angles Φ and Ψ were systematically varied in 45° steps; and the torsional potential energy was computed for each combination of Φ and Ψ.*

the torsional potential energy resulting from the actual torsional angles Φ and Ψ.

Several network architectures were tested that differed from each other with respect to number and size of hidden layers. The standard error backpropagation technique[5] was employed to adapt the weight factors ("synaptic strengths") determining the response of the network on a distinct vector of input values. In an initial phase, the whole range of torsional angles [0 to 360°] was considered; in a second phase, "fine-tuning" of the networks was done for regions in the Φ,Ψ plane energetically not too distant from the global minimum geometry.

As a general strategy, we used the data stemming from the first round of *ab initio* computations, i.e., at angles starting at 0°, as training data. The data from the second round of variations of Φ and Ψ, i.e., at angles starting at 22.5°, were not included into the training data set but served as test data in order to evaluate the predictive capabilities of the network under investigation.

RESULTS

Whole range of the Φ/Ψ-conformational space

In a first attempt, neural networks were trained to represent the whole torsional potential energy surface with Φ and Ψ varying from 0 to 360° (see Figures 1c and 2a). The training data set consisted of 81 records. For computational reasons, Φ and Ψ values were divided by 360 in order to transform the input values linearly to cover the range [0.0, 1.0]. Similarly, the training output values were transformed onto the same range of values by dividing the computed energies by the maximum energy. Several network architectures were tested; a network with one hidden layer comprising three neurons was chosen for further analysis because after 4 300 training cycles it yielded the smallest cumulative training error. Figure 2 shows the results obtained with this network: the network response (Figure 2b) represents the training data (Figure 2a) remarkably well, although there are deficiencies. In particular, the surface produced does not correctly show the expected periodic behavior with respect to variation of the torsional angles.

Notably, the total range of the torsional angles Φ and Ψ is not chemically important: conformations with torsional potential energies of several hundreds of kilocalories per mole above the global minimum are not realistically accessible for the molecule at moderate temperatures. Therefore, the performance of the network in the region of the conformational space where the *ab initio* calculation yielded torsional potential energies less than 20 kcal/mol above the global minimum as inspected in more detail. Figure 2c shows the behavior of the training data in this region, and Figure 2d shows the energy surface produced by the neural network: obviously, in the vicinity of the minimum geometry the network prediction deviates significantly from the training data set. In view of the fact, however, that the network was trained using the whole range of Φ and Ψ

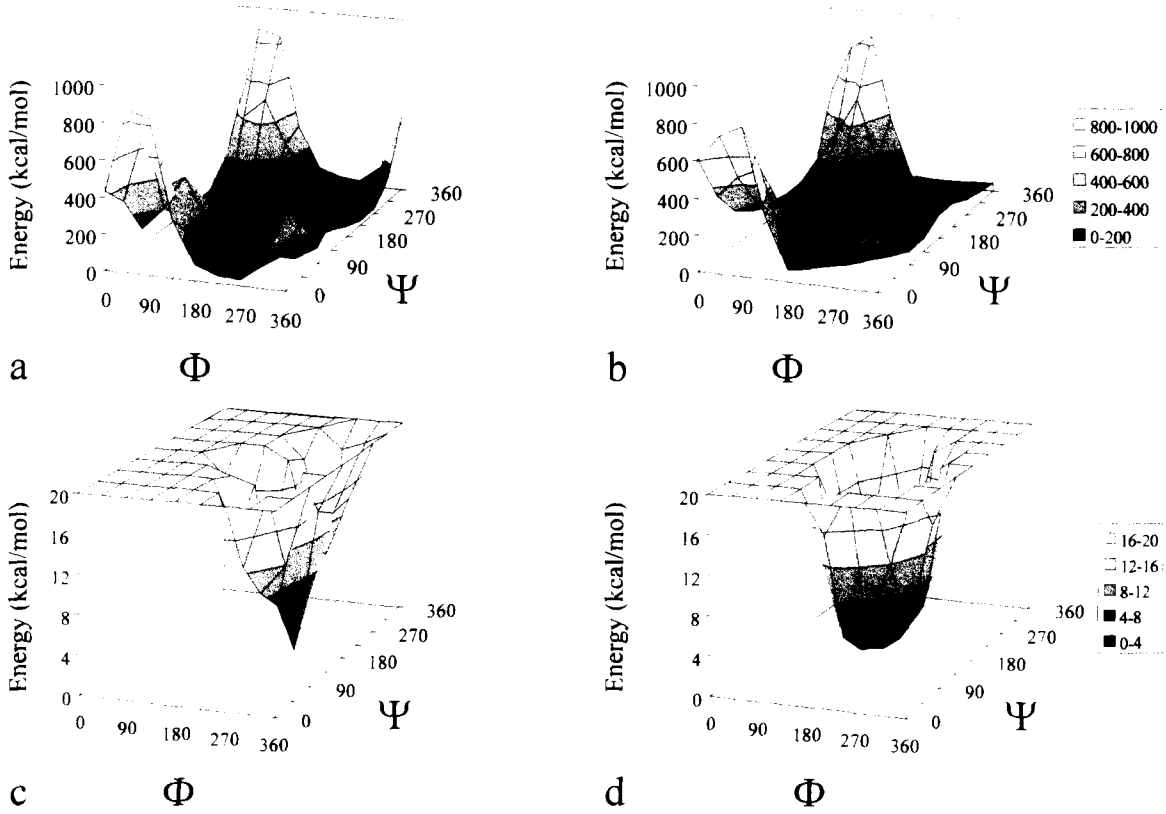

Figure 2. Torsional potential energy surfaces obtained with a neural network trained on the total range of Φ,Ψ. (a) Ab initio *results (i.e., the training data); (b) neural network results; (c)* ab initio *results in the vicinity of the global minimum conformation; (d) neural network results in the vicinity of the global minimum conformation.*

spanning a range of more than 800 kcal/mol, this failure of the network is not too surprising.

Figure 3 shows the correlation between *ab initio* torsional energies and torsional energies computed by the trained network, both for the training data (Figure 3a) and the test data (Figure 3b). While the linear correlation coefficients were apparently satisfactory (r = 0.92 for the training data, r = 0.91 for the test data), the absolute standard errors of estimated values were inacceptably high (SE_y = 79.7 kcal/mol for the training data, SE_y = 71.9 kcal/mol for the test data). As Figure 3 shows, there were several gross outliers deviating from the *ab initio* values by several hundreds of kilocalories per mole.

Selected range of the Φ/Ψ-conformational space

In a second attempt, networks were trained with the aim to better reproduce the behavior of the torsional potential in the chemically interesting region of energetically accessible conformations. For this purpose, only those input data records were included in the training set that were associated with a torsional potential energy in the range of less than and up to 20 kcal/mol above the global minimum. This criterion reduced the number of input data records to 20. The respective ranges of the two torsional angles are as follows: Φ = [135°, 315°] and Ψ = [45°, 315°]. Consequently, the following transformation rules were applied to the torsional angles to yield input values for the network: Φ′ = (Φ − 135)/180, and Ψ′ = (Ψ − 45)/270. The respective output values were obtained by dividing the torsional energies by a factor of 20. After several attempts, a 2-3-3-1 network topology (containing two hidden layers with three neurons each) was found to yield optimum results with respect to the training data.

Figure 4 shows, as contour plots, the results obtained with this network after 32 500 cycles (Figure 4b) as compared with the *ab initio* data (Figure 4a). Obviously the network is able to predict nicely the functional form of the energy surface in the investigated region.

Included in Figure 4 are scatter plots showing the correlations between *ab initio* data and network predictions for the training data (Figure 4c) and the test data (Figure 4d). Linear correlation coefficients were r = 0.98 for the training data and r = 0.49 for the test data; importantly, the standard errors of estimated values were dramatically smaller than for the network having been trained on the whole range of torsional angles: SE_y = 1.1 kcal/mol for the training data, SE_y = 5.6 kcal/mol for the test data. Notably, this network was able to predict exactly the molecular conformation with minimum energy.

The problem of overtraining

The results above were obtained by a 2-3-3-1 network after 32 500 training cycles, i.e., after training the network with the aim of optimally representing the training data. It is well known that error backpropagation networks gain increased prediction power for the training data when the number of training cycles is high but in such case they are liable to lose predictive power for previously unknown test data. To detect possible overtraining of the network used here we performed an analysis of the predictive power of the network in dependence on the number of training cycles. Figure 5a demonstrates that an optimum predictive power for the test data is already obtained after 2 000 training cycles; increase in training cycles above this limit leads, as expected, to further increase in the predictive accuracy for the training data but the prediction of test data is considerably worsened. Figure 5b (training data) and Figure 5c (test data) show the correlation between *ab initio* data and values predicted by the network when only 2 000 training cycles are used: linear correlation coefficient (r = 0.92) and standard error of estimations (SE_y = 1.8 kcal/mol) are slightly worse for the training data; however, prediction of test data is markedly improved (r = 0.69, SE_y = 3.5 kcal/mol).

Unrelaxed versus relaxed energy profiles

As mentioned in Methods, variations of molecular energies in dependence on variations of a selected set of structural variables of a given molecule can be calculated in two ways: either all other structural variables are frozen (unrelaxed energy behavior, computationally less expensive, physically incorrect) or are allowed to relax (relaxed energy behavior, computationally expensive, physically correct).

To demonstrate that neural networks trained on input data describing several (potentially many) varying structural variables of a molecule can be exploited to better approximate relaxed energy profiles/surfaces, the 2-3-3-1 network described above was used to compute the one-dimensional energy profile of tetrahydrobiopterin by systematic variation of one torsional angle, Ψ. The computation was done in

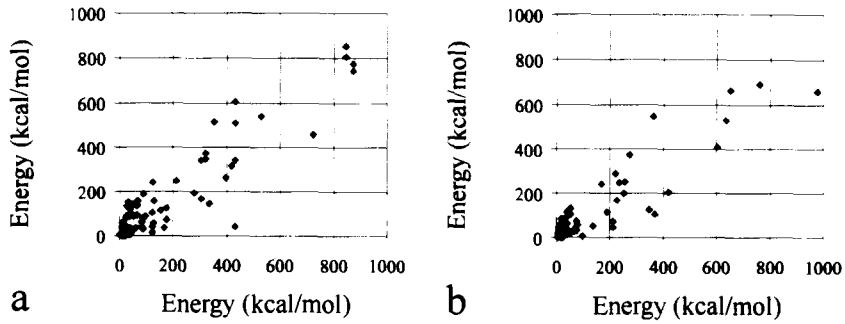

Figure 3. Correlation between ab initio *torsional potential energies (abscissa) and results of a neural network (ordinate) trained on the total range of Φ,Ψ. (a) Training data; (b) test data.*

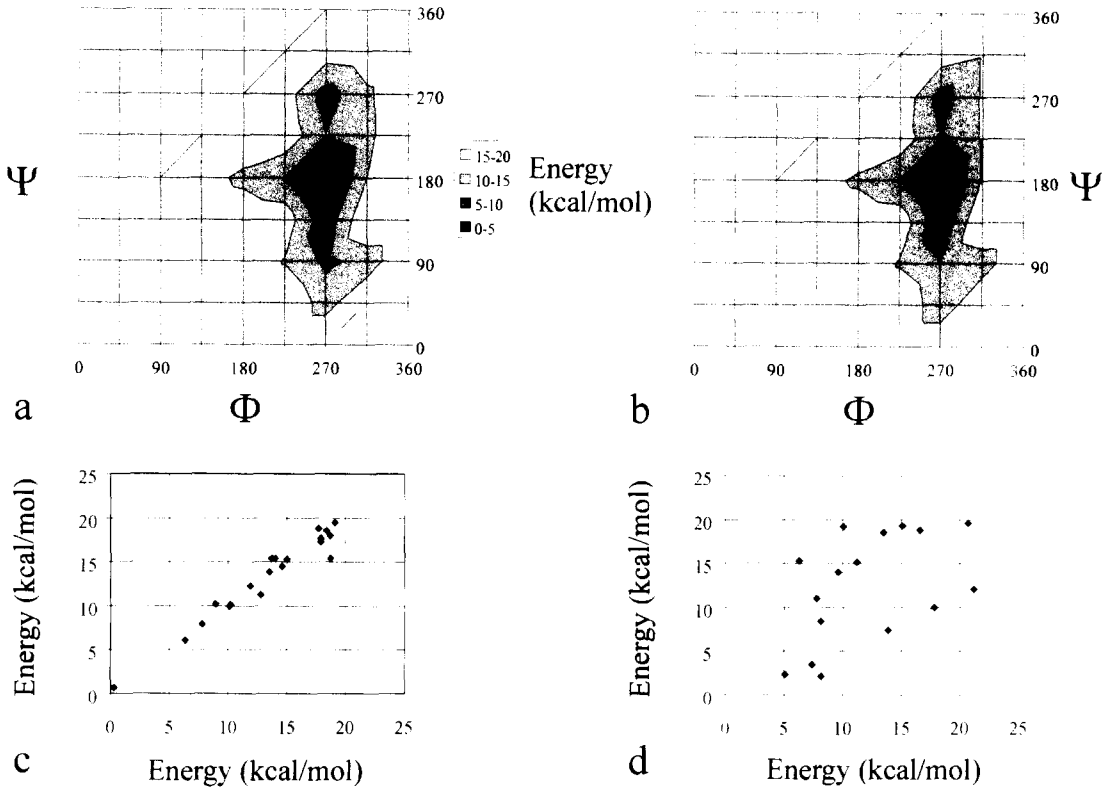

Figure 4. Torsional potential energies obtained with a neural network trained on a restricted range of Φ,Ψ *(only data records with* ab initio *torsional potential energies less than and up to 20 kcal/mol above the global minimum included). (a)* Ab initio *results (i.e., the training data); (b) neural network results; (c) correlation between* ab initio *torsional potential energies (abscissa) and results of the neural network (ordinate) for training data; (d) correlation between* ab initio *torsional potential energies (abscissa) and results of the neural network (ordinate) for test data. The results are shown for a network obtained after 32 500 training cycles.*

the following way: the 2-3-3-1 network was coded into a spreadsheet program (Excel 5.0, Microsoft, Redmond, WA), the input variable corresponding to Ψ was systematically varied, and the output of the network (corresponding to the torsional energy) was computed in either of two ways: (1) the second input value (corresponding to torsional angle Φ) was held fixed (dashed energy profile in Figure 6a; straight dashed path in Figure 6b), or (2) for each Ψ value, the built-in optimization function of the spreadsheet program (Solver) was used to search for the optimum value of Φ, given Ψ (any other optimization procedure would have served the same purpose). Geometrically, this procedure corresponds to searching, for each Ψ given, for the lowest point in the energy surface along a horizontal line. Not surprisingly, the energy profile calculated by this method (solid energy profile in Figure 6a; solid nonstraight path in Figure 6b) invariably lies below the unrelaxed profile.

DISCUSSION

Our study shows that simple feed-forward neural networks using the error backpropagation learning technique can be trained in order to provide a compact representation of a complicated molecular potential energy surface, employing training data generated by, e.g., *ab initio* quantum chemical computations. When properly done, the trained networks incorporate a reliable approximation to potential energy surfaces of the quality of the quantum chemical approach used. The networks obtained are useful only if they possess the ability to generalize, i.e., on new input vectors corresponding to new geometric conformations of the molecule they should respond with an output providing a reasonable estimate of the actual potential energy corresponding to the new geometry. Once a network is trained, its response to a certain input vector is calculated extremely rapidly, particularly in comparison to a full quantum chemical calculation at the new geometry. Of course, not only torsional potential energy could serve as output of interest; the network could, in principle, be trained to yield any other molecular property that can be calculated by quantum chemical techniques.

As our study was aimed at exploring the ability of neural networks to serve, in principle, as compact storages of quantum chemically obtained molecular information, we have kept the dimensionality of our problem low: only two torsional angles were varied. In practical use, of course, one would like to have neural networks encoding more structural information on the molecule under investigation. For example, in molecules with several rotatable single bonds information would be desirable on the effect of variations of any of the associated torsional angles, together with variations of bond lengths and bond angles. Our pilot study suggests a strategy for such considerably more demanding tasks: it seems desirable to restrict the training data to those regions in the (potentially) multidimensional space of con-

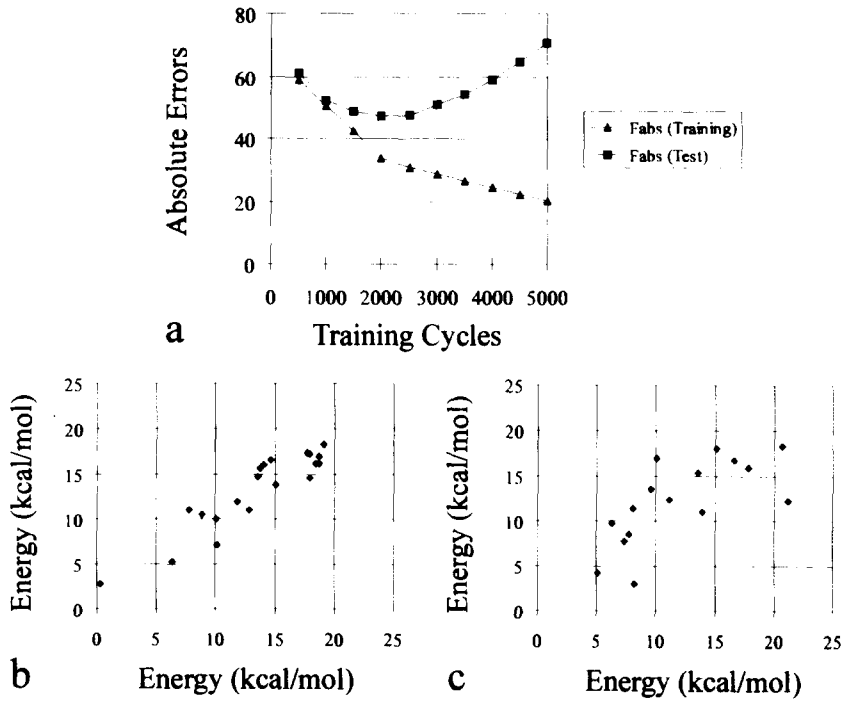

Figure 5. The problem of overtraining. (a) Validation of the predictive abilities of a neural network by comparing the total prediction error (Fabs) for the training data and the test data in dependence on the number of training cycles; (b) correlation between ab initio *torsional potential energies (abscissa) and results of the neural network (ordinate) for training data, obtained after 2 000 training cycles; (c) correlation between* ab initio *torsional potential energies (abscissa) and results of the neural network (ordinate) for test data, again after 2 000 training cycles.*

formations that are energetically accessible to the molecule. In the studied example with a two-dimensional problem it was rather easy to define such an accessible subspace simply by restricting the analysis to those ranges of Φ and Ψ where the computed energies were smaller than a given limit. How could one proceed in a more general multidimensional situation? We suggest that energetically realistic geometries of the studied molecule can be obtained by running a molecular dynamics simulation at realistic temperatures, using a suitably parametrized force field. For example, in the case of pteridines such as tetrahydrobiopterin various force fields were found to yield results that are reasonably comparable with *ab initio* calculations (see Ref. 6; and unpublished data). One could perform a simulation of a certain length, extracting molecular geometries at many time points during the simulation. Generating in this manner a set of realistic geometries would take relatively little computing time on modern personal computers. These geometries can be expected to represent a sample of different but energetically realistic structures of the molecule, and they could serve as input geometries for single-point *ab initio* calculations. Thus, a representative training data set could be generated that then could serve to train a neural network.

As mentioned in Methods, one important application of a trained neural network could be the determination of torsional potential surfaces for selected torsional angles while allowing relaxation of the molecule with respect to all other structural variables. How this could be done was demonstrated in our pilot study for the very simple case of determining the energy profile due to variation of one torsional angle while relaxing the second (Figure 6). The computational savings of such a procedure as compared with the costs for full *ab initio* treatment are obviously overwhelming. This is true even if up to several hundreds of training data were to be generated beforehand in order to obtain a network that well covers the chemically interesting regions of the structural space accessible to the molecule.

Our study might be criticized because of the relatively low number of training data records (20 in the case of the Φ,Ψ-restricted situation). In fact, however, such a sparse availability of training data is quite realistic: if a reasonably sized flexible molecule should be studied by means of the combination of *ab initio* methods and neural networks suggested here, the number of internal coordinates (bond lengths, bond angles, and torsional angles) might also be on the order of tens or even exceeding 100. Realistically one could generate a few hundreds of training data employing *ab initio* calculations; then, in comparison to the structural variables the number of training data is also relatively sparse. It seems quite promising that our study demonstrates a reliable predictive ability of the neural networks even with such a low number of training data.

We believe that neural networks could be of great help in coding quantum chemically obtained information in a compact manner. We stress, however, that a verification of the predictive abilities of the network for input vectors that are different from those used during training is absolutely indispensable. As our cross-validation example (Figure 5a) demonstrates (and as has long been known) error backpropagation networks are sensitive to the problem of overtraining. Therefore, the absolute error in predicting the training data is not sufficient to guide the training process.

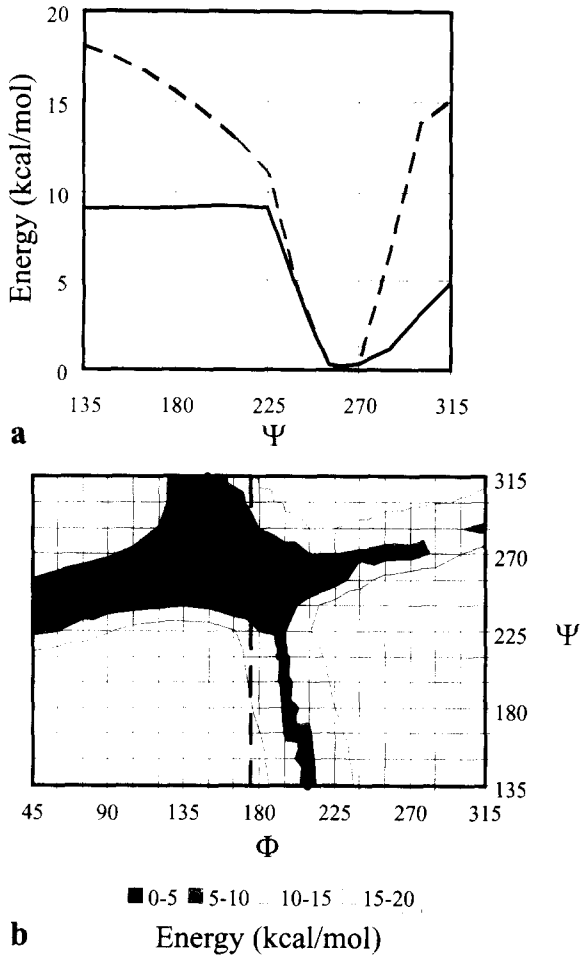

Figure 6. Relaxed vs. unrelaxed computation of the torsional potential energy in dependence on variation of torsional angle Ψ. (a) Profile of the torsional potential energy: dashed line, unrelaxed method, torsional angle Φ held fixed; solid line, relaxed method, Φ allowed to optimize. (b) The respective paths in a Φ,Ψ plane.

Rather, the quality of the network must be verified using test data unknown to the network, and the training process should be halted when optimum prediction of test data is reached.

Our work is a pilot study, but the results seem to encourage investigations of more extensive applications of neural networks as tools to store and process information obtained from quantum chemistry, particularly in problems involving molecules of biological importance, i.e., most frequently of a size rendering impossible extensive *ab initio* treatment.

REFERENCES

1 McCulloch, W.S. and Pitts, W. A logical calculus of the ideas imminent in neural nets. *Bull. Math. Biophys.* 1943, **5,** 115–133

2 Rosenblatt, F. *Principles of Neurodynamics.* Spartan, New York, 1943

3 Cooper, L.N. A possible organization of animal memory and learning. In: *Proceedings of the Nobel Symposium on Collective Properties of Physical Systems* (B. Lundquist and S. Lundquist, eds.). Academic Press, New York, 1973, pp. 252–264

4 Hopfield, J.J. Neural networks and physical systems with emergent collective computational abilities. *Proc. Natl. Acad. Sci. U.S.A.* 1982, **79,** 2554–2558

5 Rumelhart, D.E., Hinton, G.E., and Williams, R.J. Learning internal representations by error propagation. In: *Parallel Distributed Processing: Exploration in the Microstructure of Cognition* (D.E. Rumelhart and J.L. McClelland, eds.). MIT Press, Cambridge, 1986, pp. 318–364

6 Estelberger, W., Mlekusch, W., and Reibnegger, G. Conformational investigation of the cofactor (6*R*,1′*R*,2′*S*)-5,6,7,8-tetrahydrobiopterin. *Biochim. Biophys. Acta* 1995, **1249,** 23–28

7 Kaufman, S. The structure of phenylalanine hydroxylation cofactor. *Proc. Natl. Acad. Sci. U.S.A.* 1963, **50,** 1085–1093

8 Friedman, P.A., Kappelman, H., and Kaufman, S. Partial purification and characterization of tryptophan hydroxylase from rabbit hindbrains. *J. Biol. Chem.* 1972, **247,** 4165–4173

9 Tietz, A., Lindberg, M., and Kennedy, E.P. A new pteridine-requiring enzyme system for the oxidation of glyceryl ethers. *J. Biol. Chem.* 1964, **239,** 4081–4090

10 Tayeh, M.A. and Marletta, M.A. Macrophage oxidation of L-arginine to nitric oxide, nitrite and nitrate. *J. Biol. Chem.* 1989, **264,** 19654–19658

11 Werner-Felmayer, G., Werner, E.R., Fuchs, D., Hausen, A., Reibnegger, G., and Wachter, H. Tetrahydrobiopterin-dependent formation of nitrite and nitrate in murine fibroblasts. *J. Exp. Med.* 1990, **172,** 1599–1607

12 Reibnegger, G., Denny, B.J., and Wachter, H. Ab initio quantum chemical calculations on the stability of different tautomers of 6- and 7-phenacetyl pterin. *Pteridines* 1993, **4,** 23–26