

Skiebe, A. Soi, *Chem. Ber.* **1993**, *126*, 1061; c) P. J. Fagan, P. J. Krusic, D. H. Evans, S. A. Lerke, E. Johnston, *J. Am. Chem. Soc.* **1992**, *114*, 9697.

[17] Bezifferung der Fullerene-C-Atome nach P. R. Birkett, P. B. Hitchcock, H. W. Kroto, R. Taylor, *Nature* **1992**, *357*, 479.

[18] C. C. Henderson, P. A. Cahill, *Science* **1993**, *259*, 1885-1887.

[19] S. H. Hoke II, J. Molsstad, D. Dilettato, M. J. Jay, D. Carlson, B. Kahr, R. G. Cooks, *J. Org. Chem.* **1992**, *57*, 5069.

[20] a) D. A. Dixon, N. Matsuzawa, T. Fukunaga, F. N. Tebbe, *J. Phys. Chem.* **1992**, *96*, 6107; b) N. Matsuzawa, D. A. Dixon, T. Fukunaga, *ibid.* **1992**, *96*, 7594; c) N. Matsuzawa, T. Fukunaga, D. A. Dixon, *ibid.* **1992**, *96*, 10747.

[21] A. Vasella, P. Uhlmann, C. A. A. Waldraff, F. Diederich, C. Thilgen, *Angew. Chem.* **1992**, *104*, 1383; *Angew. Chem. Int. Ed. Engl.* **1992**, *31*, 1388.

[22] T. Akasaka, W. Ando, K. Kobayashi, S. Nagase, *J. Am. Chem. Soc.* **1993**, *115*, 1605.

[23] a) K. M. Creegan, J. L. Robbins, W. K. Robbins, J. M. Millar, R. D. Sherwood, P. J. Tindall, D. M. Cox, J. M. McCauley, Jr., D. R. Jones, R. Gallagher, A. B. Smith III, *J. Am. Chem. Soc.* **1992**, *114*, 1103; b) Y. Elmes, S. K. Silverman, C. Sheu, M. Kao, C. S. Foote, M. M. Alvarez, R. L. Whetten, *Angew. Chem.* **1992**, *104*, 364; *Angew. Chem. Int. Ed. Engl.* **1992**, *31*, 351.

[24] K.-D. Kampe, N. Egger, M. Vogel, *Angew. Chem.* **1993**, *105*, 1203; *Angew. Chem. Int. Ed. Engl.* **1993**, *32*, Nr. 8.

[25] a) P. J. Krusic, E. Wasserman, P. N. Keizer, J. R. Morton, K. F. Preston, *Science* **1991**, *254*, 1183; b) J. R. Morton, K. F. Preston, P. J. Krusic, S. A. Hill, E. Wasserman, *J. Phys. Chem.* **1992**, *96*, 3576; c) *J. Am. Chem. Soc.* **1992**, *114*, 5454; d) P. N. Keizer, J. R. Morton, K. F. Preston, *J. Chem. Soc. Chem. Commun.* **1992**, 1259; e) P. J. Krusic, D. C. Roe, E. Johnston, J. R. Morton, K. F. Preston, *J. Phys. Chem.* **1993**, *97*, 1736.

[26] a) P. J. Fagan, J. C. Calabrese, B. Malone, *J. Am. Chem. Soc.* **1991**, *113*, 9408; b) *Acc. Chem. Res.* **1992**, *25*, 134; c) A. L. Balch, J. W. Lee, B. C. Noll, M. M. Olmstead, *J. Am. Chem. Soc.* **1992**, *114*, 10984; d) A. L. Balch, J. W. Lee, M. M. Olmstead, *Angew. Chem.* **1992**, *104*, 1400; *Angew. Chem. Int. Ed. Engl.* **1992**, *31*, 1356; e) J. M. Hawkins, A. Meyer, T. A. Lewis, U. Bunz, R. Nunlist, G. E. Ball, T. W. Ebbesen, K. Tanigaki, *J. Am. Chem. Soc.* **1992**, *114*, 7954.

[27] a) J. H. Holloway, E. G. Hope, R. Taylor, G. J. Langley, A. G. Avent, T. J. Dennis, J. P. Hare, H. W. Kroto, D. R. M. Walton, *J. Chem. Soc. Chem. Commun.* **1991**, 966; b) H. Selig, C. Lifshitz, T. Peres, J. E. Fisher, A. R. McGhie, J. W. Romanov, J. P. McCauley, Jr., A. B. Smith III, *J. Am. Chem. Soc.* **1991**, *113*, 5475.

[28] G. A. Olah, I. Bucsi, C. Lambert, R. Anizfeld, N. J. Trivedi, D. K. Sensharma, G. K. S. Prakash, *J. Am. Chem. Soc.* **1991**, *113*, 9358.

[29] a) F. N. Tebbe, R. L. Harlow, D. B. Chase, D. L. Thorn, G. C. Campbell, J. C. Calabrese, N. Herron, R. J. Young, E. Wasserman, *Science* **1992**, *256*, 822; b) P. R. Birkett, P. B. Hitchcock, H. W. Kroto, R. Taylor, D. R. M. Walton, *Nature* **1992**, *357*, 479.

[30] A. Naim, P. B. Shevlin, *Tetrahedron Lett.* **1992**, *33*, 7097.

[31] J. W. Bausch, G. K. S. Prakash, G. A. Olah, D. S. Tse, D. C. Lorents, Y. K. Bae, R. Malhorta, *J. Am. Chem. Soc.* **1991**, *113*, 3205.

[32] a) R. E. Haufler, J. Conceicao, L. P. F. Chibante, Y. Chai, N. E. Byrne, S. Flanagan, M. M. Haley, S. C. O'Brian, C. Pan, Z. Xiao, W. E. Billups, M. A. Ciufolini, R. H. Hauge, J. L. Margrave, L. J. Wilson, R. F. Curl, R. E. Smalley, *J. Phys. Chem.* **1990**, *94*, 8634; b) C. Rüchardt, M. Gerst, J. Ebenhoch, H.-D. Beckhaus, E. E. B. Campbell, R. Tellgren, H. Schwarz, T. Weiske, S. Pitter, *Angew. Chem.* **1993**, *105*, 609; *Angew. Chem. Int. Ed. Engl.* **1993**, *32*, 584.

[33] G. A. Olah, I. Bucsi, C. Lambert, R. Anizfeld, N. J. Trivedi, D. K. Sensharma, G. K. S. Prakash, *J. Am. Chem. Soc.* **1991**, *113*, 9387.

[34] L. Y. Chiang, R. B. Upasani, J. W. Swirczewski, *J. Am. Chem. Soc.* **1992**, *114*, 10154.

[35] a) A. O. Patil, G. W. Schriver, B. Carstensen, R. D. Lundberg, *Polym. Bull.* **1993**, *30*, 187-190; b) K. E. Geckeler, A. Hirsch, *J. Am. Chem. Soc.* **1993**, *115*, 3850-3851.

[36] H. R. Karfunkel, A. Hirsch, *Angew. Chem.* **1992**, *104*, 1529; *Angew. Chem. Int. Ed. Engl.* **1992**, *31*, 1468.

[37] A. L. Balch, V. J. Catalano, J. W. Lee, M. M. Olmstead, S. R. Parkin, *J. Am. Chem. Soc.* **1991**, *113*, 8953.

[38] Siehe auch: R. Taylor, D. R. M. Walton, *Nature* **1993**, *363*, 685.

## Vorhersage der Sekundärstruktur von Proteinen aus der Aminosäuresequenz mit künstlichen neuronalen Netzen

Von *Gisbert Schneider und Paul Wrede\**

Wie faltet sich eine Aminosäuresequenz in ihre native, d. h. für die korrekte biologische Funktion des Proteins benötigte dreidimensionale Struktur? Eine umfassende Antwort auf diese Frage kann bis heute nicht gegeben werden. Zu viele Faktoren beeinflussen den Faltungsvorgang *in vivo* und *in vitro*. Ein möglicher Lösungsansatz ist die Betrachtung der räumlichen Struktur eines Proteinmoleküls als ein hierarchisch aufgebautes System (Aminosäuresequenz → Sekundärstruktur → Strukturdomänen → Tertiärstruktur und gegebenenfalls Quartärstruktur<sup>[11]</sup>) mit dem Ziel, systematisch die Teilschritte des Faltungsvorgangs bei diesem modularen Konzept zu verstehen, wobei ein erster wichtiger Schritt die Vorhersage von Sekundärstrukturen aus Aminosäuresequenzen ist.

Eine dafür häufig angewendete Methode ist das Verfahren von P. Chou und G. Fasman, das auf einer statistischen Analyse bekannter Sekundärstrukturen aufbaut<sup>[21]</sup>. Die so errechneten Präferenzen für jede Aminosäure, Teil einer Helix, einer Faltblattstruktur oder einer Schleife zu sein, können dann für die Vorhersage dieser Strukturen in neuen Sequenzen

verwendet werden. Die Vorhersage erfolgt dabei nach einem Konzept, das die Bildung einer Sekundärstruktur als ein lokales, d. h. auf einen kurzen Abschnitt der Aminosäuresequenz („window“) beschränktes Ereignis auffaßt. Ein sich über die Sequenz bewegender Ausschnitt („moving window“) wird für die Vorhersage der Sekundärstruktur an einzelnen Positionen herangezogen. Dies hat vor allem den Vorteil, daß die Sekundärstruktur an einer bestimmten Sequenzposition im Kontext zu benachbarten Aminosäuren gesehen wird. So können auch Wechselwirkungen zwischen den innerhalb des betrachteten Sequenzausschnitts liegenden Aminosäureresten mit in die Vorhersage eingehen. Es wurden viele verschiedene Konzepte zur Sekundärstrukturvorhersage entwickelt, die überwiegend auf dem Moving-window-Konzept aufbauen. Ein konsequenter Ansatz entstammt dem interdisziplinären Gebiet der molekularen Bioinformatik: Sequenzanalyse mit künstlichen neuronalen Netzen.

Neuronale Netze sind ideal für eine parallele Verarbeitung der Sequenzinformation und werden verstärkt zur Untersuchung biologischer Makromoleküle verwendet. Im Mittelpunkt steht dabei zum einen die Klassifizierung von Nucleinsäure- und Aminosäuresequenzen, zum anderen die Vorhersage von räumlichen Strukturen und deren Funktionen. Zur Klassifizierung werden überwiegend selbstorganisierende Kohonen-Netzwerke<sup>[33]</sup> und mehrschichtige Perzeptrons<sup>[44]</sup>

[\*] Priv.-Doz. Dr. P. Wrede, Dipl.-Biochem. G. Schneider  
Freie Universität Berlin  
FB Physik, AG Biophysik  
Arnimallee 14, D-14195 Berlin  
Telefax: Int. + 30/838-5186

verwendet, die eine hohe Genauigkeit und Korrektheit bei der Klassifizierung z.B. von Aminosäuresequenzen nach ihrer biologischen Funktion erzielen<sup>[5]</sup>. Von einer genauen Vorhersage der Tertiär- und Quartärstruktur aus den Aminosäuresequenzen ist man allerdings auch mit diesen Systemen noch sehr weit entfernt<sup>[6-8]</sup>. Wäre eine solche Vorhersage möglich, könnten neue Wege für ein rationales Protein-Design durch die zielgerichtete Synthese von Aminosäuresequenzen allein aufgrund theoretischer Vorüberlegungen und Modellbildung beschritten werden. Die zuverlässige Vorhersage von Sekundärstrukturen ist dabei eine wichtige Teilaufgabe. Weitere Anwendungen der unterschiedlichen Typen künstlicher neuronaler Netze in der Chemie, etwa zur Auswertung spektroskopischer Daten oder zur Kontrolle chemischer Prozesse, wurden von J. Gasteiger und J. Zupan umfassend beschrieben<sup>[9]</sup>.

Die Entwicklung künstlicher neuronaler Netze zur Vorhersage von Sekundärstrukturen aus Aminosäuresequenzen wurde erstmals 1988 von Qian und Sejnowski und 1989 von Holley und Karplus versucht<sup>[4, 6, 10]</sup>. Sie erreichten mit ihren Systemen eine Vorhersagegenauigkeit von 63–64% ( $\alpha$ ,  $\beta$  oder Random-Strukturen). Diese Genauigkeit wird mit statistischen Methoden ebenfalls erzielt. Nun ist es Burkhard Rost, Reinhard Schneider und Chris Sander vom Europäischen Molekularbiologischen Laboratorium (EMBL) in Heidelberg gelungen, bei der Sekundärstrukturvorhersage mit neuronalen Netzen eine Genauigkeit von über 70% zu erreichen<sup>[11, 12]</sup>. Auch mit Blick auf das Human Genome Project – die Sequenzierung des menschlichen Genoms – und die bereits erfolgreich abgeschlossene Sequenzierung des Chromosoms III der Hefe *S. cerevisiae*<sup>[13]</sup> ist die Weiterentwicklung und Neuentwicklung leistungsfähiger Vorhersage- und Analysemethoden für eine sinnvolle Auswertung großer Datenmengen notwendig. Die von der Gruppe am EMBL deutlich verbesserte Vorhersage von Sekundärstrukturen mit Hilfe künstlicher neuronaler Netze bei alleiniger Kenntnis der Aminosäuresequenzen zeigt neben einem möglichen Einsatz für das Design von Proteinen auch einen vielversprechenden Ansatz der molekularen Bioinformatik für die Untersuchung der gewaltigen Informationsmenge in Sequenzdatenbanken auf.

Ein künstliches neuronales Netz zur Analyse von Proteinsequenzen ist ein Computerprogramm, das Muster und Korrelationen in Sequenzdaten erkennt, wobei der Erfolg des Verfahrens unter anderem darauf beruht, daß das System relevante Information von irrelevanter Information trennen kann und sie parallel verarbeitet<sup>[4, 9]</sup>. Entsprechend dem Konzept des moving window wird durch ein künstliches neuronales Netz jede Position innerhalb des Sequenzausschnitts zu jeder anderen Position des Sequenzfensters in Beziehung gesetzt. Ein Beispiel mit einer Fenstergröße von fünf Positionen zeigt Abbildung 1.

Die von Rost et al. verwendeten Netze gehören wie die meisten zur Sequenzanalyse verwendeten Typen in die Klasse der Feedforward-Netze (Abb. 1). Der Informationsfluß läuft in eine Richtung: Von der Eingabe- zur Ausgabeeinheit des Netzes. Dies wird erreicht, indem jeder Knoten einer Netzsicht nur mit den Knoten der nächsten Schicht verbunden ist. Ein entscheidender Vorteil dieser Netzarchitektur liegt in der Fähigkeit zur Approximation beliebiger Eingabe-Ausgabe-Transformationen<sup>[14]</sup>. Das bedeutet, daß auch Transformationen zwischen Aminosäuresequenzen (Netz-

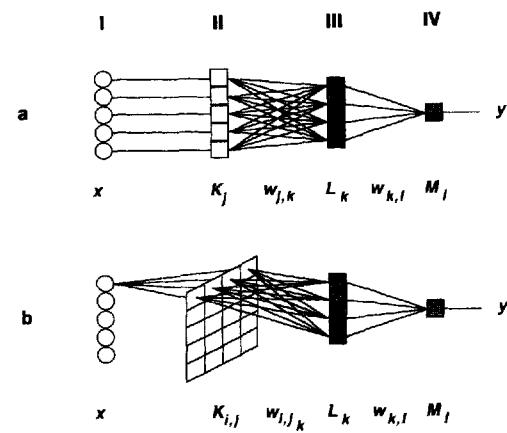


Abb. 1. Modelle einer allgemeinen Architektur dreilagiger künstlicher neuronaler Netze zur Analyse von Aminosäuresequenzen. Der Informationsfluß läuft dabei nur in eine Richtung von der Sequenz  $x$  (I) zur Ausgabeeinheit (IV), die den Wert  $y$ , z.B. einen Sekundärstrukturtyp, ausgibt („feed-forward“). Die Netzknoten der verdeckten Schicht (III) ermöglichen die Berechnung beliebiger, auch nichtlinearer Transformationsfunktionen zwischen einer Aminosäuresequenz (I) und der Ausgabeeinheit (IV) des Netzes. Eine Aminosäure  $j$  wird a) binär oder durch einen physikochemischen Parameter  $K_j$  (II) oder b) durch eine Reihe physikalisch-chemischer Eigenschaften  $K_{i,j}$  (hier  $i = 1-4$ ) (II) beschrieben.

eingabe) und deren Struktur oder Funktion (Netzausgabe) berechnet werden können. Die Transformationsfunktion  $T(x)$  eines Sequenzmusters  $x$ , die ein dreilagiges Netz (Abb. 1a) berechnen kann, ist gegeben durch die Zahl der Netzknoten ( $K_j$ ,  $L_k$ ,  $M_l$ ) und der Netzkanten mit den zugehörigen Gewichten  $w_{j,k}$ ,  $w_{k,l}$  [Gl. (a)], wobei  $S(\text{net}_{in})$ , die „Transferfunktion“ eines Netzknotens, durch Gleichung (b) gegeben

$$T(x) = S\left(\sum_i w_{i,j} S\left(\sum_{j,k} w_{j,k} \xi_j\right)\right) \quad (a)$$

$$S(\text{net}_{in}) = \frac{1}{1 + e^{-(\tau \text{net}_{in} + \sigma)}} \quad (b)$$

ist. In der Regel wird für sie eine nichtlineare sigmoide Funktion (Fermi-Funktion) gewählt;  $\tau$  bestimmt die Steilheit der Kurve,  $\sigma$  deren Wendepunkt<sup>[9]</sup>.  $\xi_j$  in Gleichung (a) ist die für die Aminosäure  $j$  gewählte Repräsentation, z.B. als zwanzigstellige Binärzahl (verteilte Codierung, Abb. 1a) oder durch einen Satz physikalisch-chemischer Parameter (Abb. 1b). Ein Vorteil der Binärcodierung liegt darin, daß eine Aminosäure formal eindeutig und vollständig beschrieben werden kann, was mit einer inhaltlichen Beschreibung durch physikalisch-chemische Eigenschaften nicht ohne weiteres möglich ist. Jedoch lassen sich Verwandtschaften und Ähnlichkeiten zwischen Aminosäuren durch deren Eigenschaften ausdrücken, was insbesondere zur Untersuchung von Sequenzen, die in der formalen Beschreibung keine oder nur eine geringe Sequenzhomologie aufweisen, von Vorteil ist. Durch die Verwendung einer nichtlinearen Transferfunktion  $S(x)$  der Netzknoten lassen sich auch nichtlineare Zusammenhänge zwischen Eingabe und Ausgabe des Netzes darstellen, die mit statistischen Verfahren, zu denen auch die erwähnte Vorhersagemethode für Sekundärstrukturen von Chou und Fasman zählt<sup>[12]</sup>, oft sehr schwierig zu berechnen sind<sup>[9]</sup>.

Wenn es also gelingt, eine Beschreibung des Zusammenhangs zwischen Sequenz und Struktur oder Funktion mit einem künstlichen neuronalen Netz zu finden, können gleichartige Strukturen oder Funktionen in neuen Sequen-

zen genau vorhergesagt werden. Eine eingehende, leicht verständliche Beschreibung der Funktionsweise und Architektur künstlicher neuronaler Netze bietet z.B. das Buch von J. Hertz, A. Krogh und R. G. Palmer<sup>[15]</sup>.

Wie „lernet“ ein künstliches neuronales Netz die richtige Transformation zwischen Sequenz und zugehöriger Struktur oder Funktion? Um eine gewünschte Transformationsfunktion – etwa für die Vorhersage von Sekundärstrukturen – zu erstellen, müssen dem Netz bekannte Beispieldaten vorliegen. Ein solcher Trainingssatz enthält Sequenzen mit bekannten, experimentell abgesicherten Sekundärstrukturen. Die freien Parameter der Transformationsfunktion des Netzes, insbesondere die Kantengewichte, werden nun mit Hilfe von Lern- und Optimierungsalgorithmen so lange zielgerichtet verändert, bis sich die vom Netz vorhergesagten Sekundärstrukturen der Aminosäuresequenzen des Trainingsatzes mit den experimentell bekannten Daten möglichst gut decken. Als Optimierungsverfahren für die Kantengewichte haben sich vornehmlich der Backpropagation-Algorithmus<sup>[16]</sup> – ein Gradientenverfahren für mehrschichtige Netzmodelle – und evolutionäre Algorithmen<sup>[17, 18]</sup> bewährt. Letztere basieren auf einer lokalen stochastischen Suche.

Ob ein „trainiertes“ Netz dann tatsächlich allgemein für die Vorhersage z.B. von Sekundärstrukturen geeignet ist, wird dadurch geklärt, daß man mit ihm Vorhersagen an einem oder mehreren Testsätzen durchführt. Ein Testsatz enthält, wie auch der Trainingssatz, Sequenzen mit bekannter Struktur, jedoch wurden die Testsequenzen nicht in der Lernphase verwendet. Ein Testsatz ist somit eine von den Trainingssequenzen unabhängige Datensammlung, mit der die Generalisierungsfähigkeit eines künstlichen neuronalen Netzes getestet werden kann. Diese Fähigkeit wird außer von der Tatsache, ob das Vorhersageproblem überhaupt mit der gewählten Methode zu lösen ist, auch von mehreren, meist empirischen Parametern bestimmt, z.B. von der gewählten Netzarchitektur, dem Lernalgorithmus, der Größe und Zahl der Netzsichten sowie einer geeigneten Datenrepräsentation. Zur computergestützten Optimierung dieser Parameter wurde bereits ein neuer Ansatz vorgestellt („Strukturevolution“)<sup>[19]</sup>, dessen Eignung für die Entwicklung von künstlichen neuronalen Netzen zur Sequenzanalyse allerdings noch gezeigt werden muß.

Entscheidend für den Erfolg einer Vorhersagemethode ist eine geeignete Darstellung der Sequenzdaten, anhand derer ein generalisierendes Merkmal für die Vorhersage gefunden werden soll. Zum einen ist auf die Repräsentation der Aminosäuren zu achten, zum anderen ist auch die Auswahl der Sequenzen und deren Aufteilung in einen oder mehrere Trainingssätze wichtig<sup>[10, 20, 21]</sup>. Besonders letzteres scheint von großer Bedeutung zu sein, wie auch Rost und Sander zeigen konnten. Sie erstellten eine Reihe von Trainingssätzen, von denen jeder eine Anzahl homologer Sequenzen enthielt, also eine „Proteinfamilie“ darstellte. Die künstlichen neuronalen Netze konnten somit in der Lernphase auf zusätzliche Infor-

mationen über die evolutionäre Verwandtschaft zwischen den Sequenzen einer „Proteinfamilie“ und der zugehörigen Sekundärstruktur zugreifen. Diese Vorgehensweise war der entscheidende Schritt gewesen, um die Genauigkeit der Sekundärstrukturvorhersage deutlich zu verbessern. Sie wird von den Autoren auch für die Entwicklung anderer Vorhersagesysteme empfohlen. Neben einer geeigneten Methode zur Merkmalsextraktion (künstliche neuronale Netze) ist eine geeignete Datenrepräsentation sehr wichtig für den Erfolg bei der Entwicklung von Vorhersagesystemen!

Da die neue Methode zur Sekundärstrukturvorhersage einem breiten Kreis von Anwendern über Elektronische Mail am EMBL zugänglich gemacht wird, kann die Generalisierungsfähigkeit des Systems („PhD-Methode“; *Profile network from Heidelberg*) tatsächlich einem harten Test unterzogen werden. Hierzu können detaillierte Anweisungen durch Absenden des Befehls „help“ an die Adresse *PredictProtein@Embl-Heidelberg.de* angefordert werden<sup>[11, 12]</sup>.

Kritisch zur Sekundärstrukturvorhersage an sich ist anzumerken, daß ihre Aussagekraft limitiert ist, auch wenn sie vollkommen korrekt ist! Entscheidend für die Funktion von Proteinen ist die Tertiärstruktur, und von deren Vorhersage ist man immer noch sehr weit entfernt. Doch zeigt der Einsatz künstlicher neuronaler Netze einen vielversprechenden Weg auf, diesem Ziel näher zu kommen, nicht zuletzt wegen der Einfachheit und der universellen Anwendbarkeit dieser Systeme.

- [1] C. Branden, J. Tooze, *Introduction to Protein Structure*, Garland Publishing, New York, 1991.
- [2] P. Y. Chou, G. D. Fasman, *Biochemistry* **1974**, *13*, 211–222, 222–245.
- [3] T. Kohonen, *Self-Organization and Associative Memory*, Springer, Berlin, 1989.
- [4] J. D. Hirst, M. J. E. Sternberg, *Biochemistry* **1992**, *31*, 7211–7218.
- [5] C. Wu, G. Whitson, J. McLarty, A. Ermongkonchai, T. C. Chang, *Protein Sci.* **1992**, *1*, 667–677.
- [6] L. H. Holley, M. Karplus, *Proc. Natl. Acad. Sci. USA* **1989**, *86*, 152–156.
- [7] H. Bohr, J. Bohr, S. Brunak, R. M. J. Cotterill, B. Lautrup, L. Norskov, O. H. Olsen, S. B. Petersen, *FEBS Lett.* **1988**, *241*, 223–228.
- [8] H. Bohr, J. Bohr, S. Brunak, R. M. J. Cotterill, H. Fredholm, B. Lautrup, S. B. Petersen, *FEBS Lett.* **1990**, *261*, 43–46.
- [9] J. Gasteiger, J. Zupan, *Angew. Chem.* **1993**, *105*, 510–536; *Angew. Chem. Int. Ed. Engl.* **1993**, *32*, 503–527.
- [10] N. Quian, T. J. Sejnowski, *J. Mol. Biol.* **1988**, *202*, 865–884.
- [11] B. Rost, C. Sander, *Nature* **1992**, *360*, 540.
- [12] B. Rost, R. Schneider, C. Sander, *Trends Biochem. Sci.* **1993**, *18*, 120–123.
- [13] S. G. Oliver und etwa 90 weitere Autoren, *Nature* **1992**, *357*, 38–46.
- [14] K. Hornik, M. Stinchcombe, H. White, *Neural Networks* **1989**, *2*, 359–366.
- [15] J. Hertz, A. Krogh, R. G. Palmer, *Introduction to the Theory of Neural Computation*, Addison-Wesley, Redwood City, CA, USA, 1991.
- [16] D. E. Rumelhart, G. E. Hinton, R. J. Williams, *Nature* **1986**, *323*, 533–536.
- [17] I. Rechenberg, *Evolutionsstrategie – Optimierung technischer Systeme nach Prinzipien der biologischen Evolution*, Frommann-Holzboog, Stuttgart, 1973.
- [18] D. E. Goldberg, *Genetic Algorithms in Search, Optimization and Machine Learning*, Addison-Wesley, Reading, MA, USA, 1989.
- [19] R. Lohmann, *Structure Evolution in Dynamic, Genetic and Chaotic Programming* (Hrsg.: B. Svoček & The IRIS Group), Wiley, New York, 1992, S. 395–411.
- [20] G. Schneider, P. Wrede, *Endocytobiosis Cell Res.* **1992**, *9*, 1–12.
- [21] G. Schneider, P. Wrede, *J. Mol. Evol.* **1993**, *36*, 586–595.