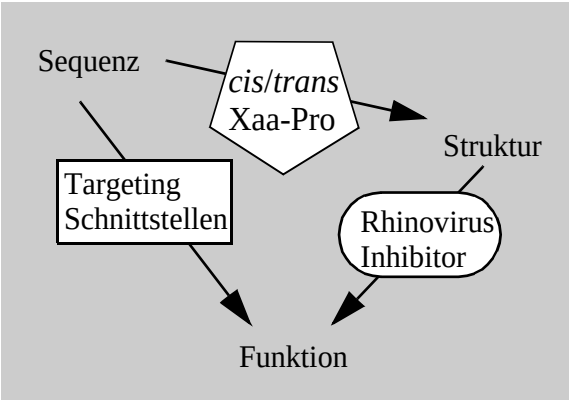
Bernd Alfred Jagla

Funktionen und Strukturen abgeleitet aus Aminosäuresequenzen mit bioinformatischen Methoden



Funktionen und Strukturen abgeleitet aus Aminosäuresequenzen mit bioinformatischen Methoden

Dissertation zur Erlangung des Doktorgrades des
Fachbereichs Biologie, Chemie, Pharmazie
der Freien Universität Berlin

vorgelegt von
Bernd Alfred Jagla
aus Berlin
1999

Diese Arbeit wurde in der Zeit von Januar 1997 bis April 1999 am Institut für Chemie der Freien Universität Berlin und am Institut für Medizinisch Technische Physik und Lasermedizin des Fachbereichs Humanmedizin der Freien Universität Berlin angefertigt. Der Verfasser versichert, die Arbeit eingenständig durchgeführt und alle Hilfsmittel angegeben zu haben.

Gutachter: Prof. Dr. Paul Wrede
Prof. Dr. Hanspeter Herzel

Tag der Disputation: 24.9.1999

Danksagungen

Mein besonderer Dank gilt Herrn Prof. Dr. Paul Wrede für seine wohlwollende Unterstützung und Betreuung bei dieser Arbeit. Herrn Prof. Dr. Ing. Gerhard Müller, Prof. h.c. Dr. h.c., der die Arbeit möglich machte, möchte ich ebenso danken.

Herrn Dr. Johannes Schuchhardt möchte ich herzlich danken für die Betreuung und die stete geduldige Hilfsbereitschaft. Herrn Dr. Gisbert Schneider danke ich für die Unterstützung während der ersten Phase der Doktorarbeit.

Herrn Prof. Dr. Volker Erdmann danke ich für seine idelle und materielle Unterstützung beim Zustandekommen und der Durchführung dieser Arbeit.

Herrn Stefan Müller und Herrn Dr. Arif Malik danke ich für die Zusammenarbeit in der AG Bioinformatik.

Ich danke des weiteren den Herren Prof. Dr. Heinz Zeichhardt und Herrn Dr. Hans-Peter Grunert für das Testen der nona-Peptide sowie für die Unterstützung meiner Bemühungen.

Herrn Prof. Dr. Reinhart Heinrich danke ich stellvertretend für das gesamte Graduiertenkolleg "Dynamik und Evolution Makromolekularer und zellulärer Prozesse", an dem ich teilhaben durfte.

Dem ZIB, insbesondere Herrn Dr. Manfred Stolle danke ich für die Betreuung meiner Projekte an der CRAY.

Herrn Dr. Dieter Riedel danke ich für Unterstützung bei der Benutzung der Rechner der ZEDAT.

Darüber hinaus bedanke ich mich bei allen, die mich bei der Durchführung dieser Arbeit tatkräftig unterstützt haben.

Abkürzungsverzeichnis

Å	Angström (=10^{-10} m)
3D	dreidimensional
ACN	Adaptiv kodierendes KNN
C	Kohlenstoffatom
CATH	class(C), architecture(A), topology(T) and homologous superfamily (H)
CsA	Cyclosporin A
Cyp	Cyclophilin
D_2O	Schweres Wasser
DSSP	Verzeichnis von Protein Sekundärstrukturen
ER	Endoplasmatisches Retikulum
FK506	Immunsuppressivum
FKBP	FK506 bindendes Protein
H	Wasserstoffatom
H_2O	Wassermolekül
HRV	Humanes Rhinovirus
i	Index i
K, $\vec{k}$	Eingabezahl, -vektor
kD	Kilo Dalton
KNN	Künstliches Neuronales Netz
λ	Lernrate in einem evolutionären Algorithmus
μ	Index der Sequenzen
MHC	Major Histocompatibility Complex
n	Index der Kodierungsvektoren
N	Stickstoffatom
N_c	Anzahl der Kodierungsvektoren
N_m	Anzahl der Muster
NMR	Kernmagnetresonanz
NN	Neuronales Netz
N_s	Anzahl der Aminosäuren in einer Sequenz
O	Sauerstoffatom
PDB	Protein Datenbank
PPIase	Peptidylprolylisomerasen
θ	Schwellenwert (threshold)
σ	Lernschrittweite
s	Standardabweichung
Σ	Summenfunktion
s^2	Varianz
SAR	Sequenz-Aktivitäts-Relation
SME	Simulierte Molekulare Evolution
SOK	Selbst Organisierende Karte
SPase	Signalpeptidase
SPC18, SPC21, SPC22/23, GP23, P19, SPC12, Sec11p	Signalpeptidaseuntereinheiten
SRP	Signalerkennungspartikel
w, $\vec{w}$	Gewichte
WWW	World Wide Web
X, Xaa	eine der 20 natürlichen Aminosäuren
x, y	beliebige reelle Zahl

Tabelle der hier verwendeten Aminosäuren im Ein- und Drei-Buchstabenkode

A	Ala	Alanin
C	Cys	Cystein
D	Asp	Asparaginsäure
E	Glu	Glutaminsäure
F	Phe	Phenylalanin
G	Gly	Glycin
H	His	Histidin
I	Ile	Isoleucin
K	Lys	Lysin
L	Leu	Leucin
M	Met	Methionin
N	Asn	Asparagin
P	Pro	Prolin
Q	Gln	Glutamin
R	Arg	Arginin
S	Ser	Serin
T	Thr	Threonin
V	Val	Valin
W	Trp	Tryptophan
Y	Tyr	Tyrosin

Weiterhin wurden allgemein übliche Abkürzungen verwendet.

Inhaltsverzeichnis

KAPITEL 6 Zusammenfassung

Verschiede Methoden zur Funktions- und Strukturvorhersage ausgehend von der Primärstruktur wurden vorgestellt.

Ein adaptiv kodierendes Neuronales Netz (ACN) ist entwickelt worden, mit dem es möglich ist, die charakteristischen physikochemischen Eigenschaften in Proteinsequenzen zu bestimmen. Hiermit ist es möglich, 96% der experimentell verifizierten Schnittstellen von humanen sekretorischen Proteinen vorherzusagen. Bisher waren nur Vorhersagen mit einer Genauigkeit von bis zu 68% möglich. Außerdem können mit den ACN Mutationen in den Schnittstellenregionen humaner sekretorischer Proteine korrekt vorhergesagt werden. Es wurde somit erfolgreich gezeigt, wie aus der Sequenzinformation auf eine Funktion geschlossen werden kann.

Die Vorhersage der Konformation von Peptidylprolylbindungen ist mit dem vorliegenden Datensatz nicht zufriedenstellend lösbar. Für eine erfolgversprechende Analyse müßte die 10- bis 20-fache Menge an Daten vorliegen. Trotzdem konnte gezeigt werden, daß die MHC-I Proteine eine charakteristische *cis*-Prolingruppe besitzen, die eine Unterscheidung von anderen Proteinen erlaubt.

Die Simulierte Molekulare Evolution wurde erfolgreich zum Entwurf zellprotektiver Peptide angewendet. Ein Algorithmus zum Design inhibitorischer Proteine, basierend auf der Analyse von Protein-Protein-Wechselwirkungen, wurde entwickelt und erfolgreich auf zellprotektive Experimente getestet.

KAPITEL 7 Summary

Serveral methods were presented for the prediction of function and structure based on the primary structure of proteins.

An adaptive coding artificial neural network (ACN) was developed that characterizes the physico chemical features in protein sequences. It is now possible to identify 96% of experimentally verified cleavage sites in human secretory proteins. Hitherto only 68% of an independent test data were correctly identified. Furthermore mutations within the cleavage site region of human secretory proteins are correctly predicted. It is shown how to use ACN to build up a sequence activity relation.

It was not possible to predict the conformation of the peptidyl prolyl bond using the available data. However there should be 10 - 20 times more non-homologues sequence data to make reliable predictions. Nevertheless, it was possible to show that MHC-I proteins have a characteristic *cis*-proline group that can be used to distinguish those proteins from others.

The simulated molecular evolution was successfully applied to design cell protective peptides against the Human Rhinovirus. An algorithm for the design of inhibitory peptides using the structual information of the receptor protein has been developed. Using this method other peptides were designed. All peptides were experimentally tested, showing that theoretical and practical results are in agreement.

KAPITEL 7 Lebenslauf

Name:	Bernd Alfred Jagla
Geburtstag:	12. April 1970
Geburtsort:	Berlin
Eltern:	Eheleute Ludwig und Dr. med. Maria Jagla
Schule	
1976 - 1977	Matthias-Claudius-Schule in Düsseldorf
1977 - 1978	Städt. Kath. Grundschule Düsseldorf
1978 - 1980	Mühlenau-Grundschule Berlin, Zehlendorf
1980 - 1982	Erich-Kästner-Grundschule Berlin, Zehlendorf
1982 - 1986	Arndt-Oberschule Berlin, Zehlendorf
1986 - 1989	Walther-Rathenau-Oberschule Berlin, Wilmersdorf
23.5.1989	Abitur
Studium	
WS 1989	Aufnahme des Studiums der Chemie (Diplom) an der Freien Universität Berlin
WS 1990- SS 1991	Studium der Informatik für Lehramt als 2. Fach
25.9.1991	Vordiplom in Chemie
WS 1991	Fortführung des Chemiestudiums an der Universität Erlangen-Nürnberg
SS 1992	Fortsetzung des Studiums an der Technischen Universität München
24.10.1995	Diplom in Chemie an der TU-München (Prof. Ugi) Zusammenarbeit mit der FU-Berlin (Prof. Wrede) Titel: Sequenzbasierte Vorhersage der Konformation von prolinhaltigen Peptidgruppen in Proteinen