

Stephan Klokow

Entwicklung einer auf OpenGL basierenden
grafischen Oberfläche für das de novo
Design von Proteinstrukturen

Diplomarbeit

Bibliografische Information der Deutschen Nationalbibliothek:

Bibliografische Information der Deutschen Nationalbibliothek: Die Deutsche Bibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.d-nb.de/> abrufbar.

Dieses Werk sowie alle darin enthaltenen einzelnen Beiträge und Abbildungen sind urheberrechtlich geschützt. Jede Verwertung, die nicht ausdrücklich vom Urheberrechtsschutz zugelassen ist, bedarf der vorherigen Zustimmung des Verlanges. Das gilt insbesondere für Vervielfältigungen, Bearbeitungen, Übersetzungen, Mikroverfilmungen, Auswertungen durch Datenbanken und für die Einspeicherung und Verarbeitung in elektronische Systeme. Alle Rechte, auch die des auszugsweisen Nachdrucks, der fotomechanischen Wiedergabe (einschließlich Mikrokopie) sowie der Auswertung durch Datenbanken oder ähnliche Einrichtungen, vorbehalten.

Copyright © 2004 Diplomica Verlag GmbH
ISBN: 9783832497958

Stephan Klokow

**Entwicklung einer auf OpenGL basierenden grafischen
Oberfläche für das de novo Design von Proteinstruktu-
ren**

Stephan Klokow

Entwicklung einer auf OpenGL basierenden grafischen Oberfläche für das de novo Design von Proteinstrukturen

Diplomarbeit
Fachhochschule Merseburg
Fachbereich Informatik und angewandte Naturwissenschaften
August 2004



Diplom.de

Diplomica GmbH _____
Hermannstal 119k _____
22119 Hamburg _____

Fon: 040 / 655 99 20 _____
Fax: 040 / 655 99 222 _____

agentur@diplom.de _____
www.diplom.de _____

Stephan Klokow

Entwicklung einer auf OpenGL basierenden grafischen Oberfläche für das de novo Design von Proteinstrukturen

ISBN-10: 3-8324-9795-1

ISBN-13: 978-3-8324-9795-8

Druck Diplomica® GmbH, Hamburg, 2006

Zugl. Fachhochschule Merseburg, Merseburg, Deutschland, Diplomarbeit, 2004

Dieses Werk ist urheberrechtlich geschützt. Die dadurch begründeten Rechte, insbesondere die der Übersetzung, des Nachdrucks, des Vortrags, der Entnahme von Abbildungen und Tabellen, der Funksendung, der Mikroverfilmung oder der Vervielfältigung auf anderen Wegen und der Speicherung in Datenverarbeitungsanlagen, bleiben, auch bei nur auszugsweiser Verwertung, vorbehalten. Eine Vervielfältigung dieses Werkes oder von Teilen dieses Werkes ist auch im Einzelfall nur in den Grenzen der gesetzlichen Bestimmungen des Urheberrechtsgesetzes der Bundesrepublik Deutschland in der jeweils geltenden Fassung zulässig. Sie ist grundsätzlich vergütungspflichtig. Zuwiderhandlungen unterliegen den Strafbestimmungen des Urheberrechtes.

Die Wiedergabe von Gebrauchsnamen, Handelsnamen, Warenbezeichnungen usw. in diesem Werk berechtigt auch ohne besondere Kennzeichnung nicht zu der Annahme, dass solche Namen im Sinne der Warenzeichen- und Markenschutz-Gesetzgebung als frei zu betrachten wären und daher von jedermann benutzt werden dürften.

Die Informationen in diesem Werk wurden mit Sorgfalt erarbeitet. Dennoch können Fehler nicht vollständig ausgeschlossen werden, und die Diplomarbeiten Agentur, die Autoren oder Übersetzer übernehmen keine juristische Verantwortung oder irgendeine Haftung für evtl. verbliebene fehlerhafte Angaben und deren Folgen.

© Diplomica GmbH

<http://www.diplom.de>, Hamburg 2006

Printed in Germany

Diese Diplomarbeit wurde in Zusammenarbeit mit der

ACGT ProGenomics AG

mit Sitz in Halle (Saale) angefertigt.

Inhalt

Abbildungsverzeichnis	4
Symbolverzeichnis	7
1 Motivation und Grundlagen	9
1.1 Motivation.....	9
1.2 Das Protein.....	11
1.2.1 Aufbau von Proteinen	11
1.2.2 Faltung von Proteinen.....	13
1.2.3 Die α -Helix	13
1.2.4 Das β -Faltblatt	15
1.2.5 Das Rückgrat eines Proteins	16
1.2.6 Das Ramachandran-Diagramm.....	17
1.3 Die Tetrapeptidfunktionen der ACGT ProGenomics AG.....	19
1.4 Grundlagen der Grafikprogrammierung	21
1.4.1 OpenGL	21
1.4.1 Theoretische Grundlagen von OpenGL.....	24
1.4.2 DirectX	36
1.4.3 Das Tao-Framework	37
2 Herangehensweise.....	40
3 Programmierung.....	42
3.1 Das .NET-Framework.....	42

3.2	Softwareentwicklung	43
3.2.1	Phasen	43
3.2.2	Begriffe	44
4	Programmierung der grafischen Oberfläche	47
4.1	Erstellen der Sekundärstrukturelemente in OpenGL	47
4.2	Geometrische und topologische Eigenschaften der Sekundärstrukturelemente	59
5	Berechnen der Sequenz	62
5.1	Parameter für die Berechnung	62
5.2	Erstellen der Sequenz.....	69
5.3	Auswertung der Ergebnisse	85
6	Zusammenfassung und Ausblick.....	90
6.1	Zusammenfassung und offene Aufgaben.....	90
6.2	Ausblick.....	91
7	Literaturverzeichnis	92
8	Anhang.....	97
A)	Programmierbeispiel A	97
B)	Programmierbeispiel B	98
C)	Programmierbeispiel C	100
D)	Die energetisch besten Sequenzen für das Protein 1QYS.....	104

Abbildungsverzeichnis

Abbildung 1: HIV Protease-Inhibitor Komplex	11
Abbildung 2: Peptidbindung zwischen einem C- und einem N-Atom	12
Abbildung 3: Rechtsgängige α -Helix	14
Abbildung 4: β -Faltblatt	15
Abbildung 5: N-C α - und C α -C-Bindung	17
Abbildung 6: Sterische Bindung zwischen Carbonyl-Sauerstoff und Amid-Wasserstoff	18
Abbildung 7: Ramachandran-Diagramm	19
Abbildung 8: ψ und ϕ zwischen der dritten und der vierten Aminosäure	21
Abbildung 9: OpenGL Grafik-Pipeline	22
Abbildung 10: Punkttransformationen Translation und Rotation	25
Abbildung 11: 4*4-Matrix	26
Abbildung 12: Addition der Verschiebungskoordinaten	27
Abbildung 13: Skalierung eines Punktes	24
Abbildung 14: Rotation eines Punktes durch Multiplikation	27
Abbildung 15: Translationsgleichung mit inhomogenen Koordinaten	28
Abbildung 16: Orthogonale und perspektivische Projektion	28
Abbildung 17: Reihenfolge bei der Abarbeitung von Transformationen	30
Abbildung 18: Kamera und <i>Viewport</i> im Ursprung	31
Abbildung 19: Objekt mit <i>Gouraud-Shading</i>	33
Abbildung 20: Objekt mit <i>Flat-Shading</i>	34
Abbildung 21: Das Drahtgittermodell eines Würfels	34
Abbildung 22: Implementierung der Dlls	38
Abbildung 23: SimpleOpenGLControl	39

Abbildung 24: Klassendiagramm der entwickelten Software	46
Abbildung 25: Listenbefehle in der Übersicht	48
Abbildung 26: Quad- bzw. Triangle-Strukturen in einem Faltblatt	49
Abbildung 27: Faltblatt-Objekt im Programm	51
Abbildung 28: Helix-Objekt im Programm	51
Abbildung 29: Helix bestehend aus Quad-Objekten	52
Abbildung 30: Übersicht der GLUT-Befehle	51
Abbildung 31: Faltblatt-Objekt mit Atomen	54
Abbildung 32: Generierung der Aminosäuren	55
Abbildung 33: Generierung eines Sekundärstrukturelementes	56
Abbildung 34: Position von Faltblatt und Aminosäurekette	56
Abbildung 35: Positionen des Faltblatts und der Atome	57
Abbildung 36: Helix-Objekt im Koordinatenursprung	57
Abbildung 37: Die Generierung der AS (Atome)	58
Abbildung 38: Position von Helix und Aminosäuren	56
Abbildung 39: Positionen der Helix und der Atome der Aminosäuren stimmen jetzt überein	58
Abbildung 40: Winkelkonformation in Sekundärstrukturen	60
Abbildung 41: Verdrillungswinkel und Ausrichtung der Stränge bei antiparallelen Faltblatt	61
Abbildung 42: Verdrillungswinkel und Ausrichtung der Stränge bei parallelen Faltblatt	61
Tabelle 5.1: Ergebnisdatensatz des Tetrapeptides ACNK	62
Abbildung 43: Dichtefunktion von ACNK	63
Abbildung 44: Richardson & Richardson Nomenklatur	65
Tabelle 5.2: Anzahl der Tetrapeptide an den Positionen N_1 , I und C_1 in einer Helix	66
Tabelle 5.3: Anzahl der Tetrapeptide an den Positionen N_1 , I und C_1 in einem Faltblatt	66
Abbildung 45: Bestimmung der Wahrscheinlichkeitswerte für Tetrapeptide	67
Abbildung 46: 3-D-Plot der Dichtefunktion des Tetrapeptides IAHV	68

Abbildung 47: Hierarchische Baumstruktur	71
Abbildung 48: Struktur des <i>DataSet</i> -Modells	72
Abbildung 49: Struktur <i>NodeChild</i> im Programm	73
Abbildung 50: Grafische Darstellung der Proteinstruktur des Proteins 1G28C	75
Abbildung 51: Erstellen eines <i>View</i> auf der Datenbank für Faltblattstrukturen	77
Abbildung 52: Erstellen eines <i>View</i> auf der Datenbank für Helixstrukturen	77
Abbildung 53: Selektion der Ausgangsknoten für Helixstrukturen	78
Abbildung 54: Selektion der Ausgangsknoten für Faltblattstrukturen	78
Abbildung 55: Erstellen einer Sequenz aus Tetrapeptiden (schematisch)	80
Abbildung 56: Hydrophathie Index von Aminosäure-Seitenketten	83
Abbildung 57: Abarbeitung der Ausgangsknotenpunkte	84
Tabelle 5.4: Die Anzahl der berechneten Sequenzen für ein Faltblatt-Objekt	84
Tabelle 5.5: Die Anzahl der berechneten Sequenzen für ein Helix-Objekt	84
Abbildung 58: 3D Darstellung des Proteins 1QYS	87
Abbildung 59: Ausschnitt aus einer Ergebnisdatei nach der Kraftfeldberechnung	88
Abbildung 60: Ausschnitt aus einer Ergebnisdatei mit der aufsummierten Gesamtenergie	88
Abbildung 61: Die energetisch beste Struktur im Vergleich zur Originalstruktur	104
Abbildung 62: Die energetisch zweitbeste Struktur im Vergleich zur Originalstruktur	104
Abbildung 63: Die energetisch drittbeste Struktur im Vergleich zur Originalstruktur	104