

**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 Verlags. 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

**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

<b>Abbildungsverzeichnis .....</b>	<b>4</b>
<b>Symbolverzeichnis .....</b>	<b>7</b>
<b>1      Motivation und Grundlagen .....</b>	<b>9</b>
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 $\alpha$ -Helix .....	13
1.2.4      Das $\beta$ -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
<b>2      Herangehensweise.....</b>	<b>40</b>
<b>3      Programmierung.....</b>	<b>42</b>
3.1      Das .NET-Framework.....	42



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

# Abbildungsverzeichnis

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

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

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