

3D-ELEKTRONENMIKROSKOPIE
DREIER RESPIRATORISCHER PROTEINE
UNTER VERBESSERUNG DER
BIOINFORMATISCHEN ABLÄUFE

Dissertation

Zur Erlangung des Grades

„**Dr. rer. nat.**“

an der Johannes Gutenberg-Universität Mainz

Mario Dejung

Geb. am 31. Januar 1982 in Mainz

Mainz, 24. April 2014

Dekan: [REDACTED]

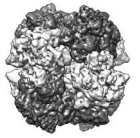
1. Berichterstatter: [REDACTED]

2. Berichterstatter: [REDACTED]

Tag der Prüfung: 17.07.2014

„The greater danger for most of us
lies not in setting our aim too high and falling short;
but in setting our aim too low, and achieving our mark.“

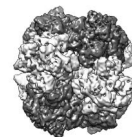
Michelangelo



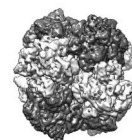
Inhaltsverzeichnis

Abkürzungsverzeichnis	V
Abbildungsverzeichnis	VII
Tabellenverzeichnis	XI
A Einleitung	1
1 Das Elektronenmikroskop als Werkzeug zur Proteinanalyse	1
2 Respiratorische Proteine	4
2.1 Das Hämoglobin des Wasserfloh <i>Daphnia pulex</i>	6
2.2 Das Hämocyanin des Flusskrebses <i>Astacus leptodactylus</i>	8
2.3 Das Hämocyanin des Pfeilschwanzes <i>Limulus polyphemus</i>	8
3 Bioinformatische Methoden in der 3D-EM	9
B Material und Methoden	11
1 Tiere	11
2 Biochemische Methoden	13
2.1 Isolierung der Proteine	13
2.2 Gelfiltrationschromatographie	14
3 Elektronenmikroskopie	15
3.1 Allgemeine Grundlagen	15
3.2 Ausrüstung	18
3.3 Probenherstellung für das Elektronenmikroskop	22
3.4 Elektronendetektion	24
4 Computer	26
4.1 Rechen-Cluster	26
4.2 Arbeitsplatzrechner	27
5 Software	27
6 Bildverarbeitende Methoden	30
6.1 Automatische Stapelverarbeitung mit Appion	33
6.2 Bildkorrektur des Partikelstapels	38
6.3 Referenzfreie Ausrichtung	38
6.4 Klassifizierung	40
6.5 Winkelzuweisung	42
6.6 3D-Rekonstruktion	44

6.7	Filtern der 3D-Rekonstruktion	44
6.8	Referenzenerstellung	47
6.9	Partikelausrichtung	47
6.10	Qualität und Auflösung der Rekonstruktion	48
6.11	Iterative Verfeinerung	50
6.12	Einpassen atomarer Modelle	50
6.13	Molekulares Modellieren mit Gromacs	51
6.14	Verwendete Module	51
7	Programmiersprachen und -methoden	53
7.1	Programmiersprachen	53
7.2	<i>Test-driven Development</i>	54
7.3	<i>Clean-Code</i>	54
C	Ergebnisse	55
1	3D-EM des multimeren Hämoglobins von <i>Daphnia pulex</i>	55
2	3D-EM des 2x6meren Hämocyanins von <i>Astacus leptodactylus</i>	60
3	3D-EM des 8x6meren Hämocyanins von <i>Limulus polyphemus</i>	64
3.1	Erstellen einer 7,1 Å-Rekonstruktion	64
3.2	Beschreibung der 7,1 Å-Rekonstruktion	68
3.3	Flexibles Fitting der Untereinheiten mit Direx	69
3.4	Korrektur der 3D-Rekonstruktion im Bereich der III A–III A-Bindestelle	70
3.5	Molekulardynamik-Simulation der III A–III A-Bindestelle	76
3.6	Dichtekarte aus MD simulieren	87
4	Entwicklung einer grafischen Benutzeroberfläche und neuer Software- Tools	91
4.1	Die grafische Benutzeroberfläche MDEMTtools	91
4.2	Die Konfigurationsdatei	92
4.3	Verwendete Programmierkonzepte	95
4.4	Verbesserung der Administration am Rechen-Cluster	95
4.5	Dokumentation der neuen Software als DokuWiki	96
4.6	Entwickelte Werkzeuge	99
D	Diskussion	105
1	3D-EM des multimeren Hämoglobins von <i>Daphnia pulex</i>	105
2	3D-EM des 2x6meren Hämocyanins von <i>Astacus leptodactylus</i>	109
3	3D-EM des 8x6meren Hämocyanins von <i>Limulus polyphemus</i>	112
3.1	MD-Simulation des <i>L. polyphemus</i> -Hämocyanin	112
3.2	Fixierung von Aminosäuren vor der MD-Simulation	113
3.3	Vergleich der 3D-Rekonstruktionen des <i>L. polyphemus</i> -Hämocyanins	116
3.4	Abmessungen des Hämocyanins	118



4	Programmierarbeiten	122
4.1	<i>Graphical user interface</i> (GUI)	122
4.2	Administration am Rechen-Cluster	123
4.3	DokuWiki	123
4.4	Entwickelte Software-Tools	124
Zusammenfassung		131
E Appendix		133
1	2%ige Uranylacetat Lösung	133
2	2%ige Uranylformiat Lösung	133
3	Negativfilme entwickeln	133
4	Pufferrezepte	134
5	Graphen der MD-Simulation	134
Literaturverzeichnis		151
Danksagung		167
Erklärung und Copyright		169



Abkürzungsverzeichnis

1D	eindimensional
2D	zweidimensional
3D	dreidimensional
FSC	Fourier-Schalen-Korrelation (<i>fourier shell correlation</i>)
GUI	grafische Benutzeroberfläche (<i>graphical user interface</i>)
CTF	Kontrast-Transfer-Funktion (<i>contrast transfer function</i>)
GB	Gigabyte
TB	Terabyte
MPI	Max-Planck-Institut
CMOS	<i>complementary metal oxide semiconductor</i>
DPI	Bildpunkte pro Inch (<i>dots per inch</i>)
CSV	kommaseparierte Werte <i>comma-separated values</i>
bash	<i>Bourne-again shell</i>
JGU	Johannes Gutenberg-Universität
CP3	Cryoplunger™3
NAS	Netzwerk-speicher (<i>network attached storage</i>)
ISAC	iterative, stabile Ausrichtung und Klassifizierung (<i>iterative stable alignment and clustering</i>)
TDD	Testgetriebene Entwicklung (<i>test-driven development</i>)
FFT	Schnelle Fourier-Transformation (<i>fast Fourier transform</i>)
PMSF	Phenylmethylsulfonylfluorid
EM	Elektronenmikroskop
3D-EM	3D-Elektronenmikroskopie
TEM	Transmissionselektronenmikroskop
DQE	Effizienz detektierbarer Menge (<i>detective quantum efficiency</i>)
z. B.	zum Beispiel
SNR	Signal-Rausch-Verhältnis (<i>signal to noise ratio</i>)
MD	Molekulardynamik (<i>molecular dynamics</i>)
PSF	Punktspreizfunktion (<i>point spread function</i>)
Hb	Hämoglobin
Hc	Hämocyanin
d. h.	das heißt
etc.	<i>et cetera</i>
bzw.	beziehungsweise
GDFR	<i>gridding-based direkt fourier reconstruction</i>

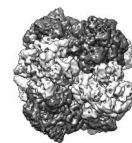
ca.	<i>circa</i>
MPI	<i>message passing interface</i>
DIN	Deutsche Industrie für Normung
TRIS	Trishydroxymethyl-aminomethan
SAXS	Kleinwinkelstreuung (<i>Small-angle X-ray scattering</i>)
ESI-MS	Elektronenspray-Ionisation Massenspektrometrie
Abb.	Abbildung

Liste der Aminosäuren

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

Einheiten

Einheit	Abkürzung	Einheit	Abkürzung
Kilometer	km	Meter	m
Centimeter	cm	Millimeter	mm
Micrometer	μm	Nanometer	nm
Ångström	Å	Picometer	pm
Inch	in		
Gram	g	Milligramm	mg
Microgramm	μg	Nanogramm	ng
Kilodalton	kDa		
Minuten	min	Sekunden	s
Microsekunden	μs	Nanosekunden	ns
Picosekunden	ps		

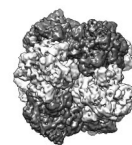


Abbildungsverzeichnis

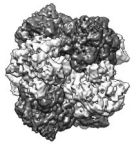
A.1	Transmissionselektronenmikroskope	2
A.2	Die Anzahl atomarer Strukturen in der PDB	3
A.3	Vergleich des aktiven Zentrums von Hämoglobin und Hämocyanin	4
A.4	Struktur einer Häm-Domäne	6
A.5	Struktur einer Untereinheit des Artrophoden-Hämocyanins	7
A.6	Struktur des 1x6meren Hämocyanins	8
A.7	Struktur des <i>Limulus polyphemus</i> -Hämocyanins aus Martin et al. (2007)	9
B.1	Der Gemeine Wasserfloh <i>Daphnia pulex</i>	11
B.2	Der Galizische Sumpfkrebs <i>Astacus leptodactylus</i>	12
B.3	Der Atlantische Pfeilschwanzkreb <i>Limulus polyphemus</i>	13
B.4	Merstufige Abbildung im Elektronenmikroskop	16
B.5	Kontrast Transfer Funktion	19
B.6	Schematische Darstellung eines C-flat Objektträgers	20
B.7	Gatan Cryoplunger™3 mit Atmosphärenkammer	21
B.8	Tecnai G ² 12 BioTWIN	23
B.9	Elektronenmikroskopische Aufnahme des <i>A. leptodactylus</i> -Hämocyanins in amorphen Eis	31
B.10	Arbeitsablaufdiagramm für 3D-Elektronenmikroskopie	32
B.11	Aufnahmenqualität und Kontaminationen	34
B.12	Partikelverteilung im vitrifizierten Eis	34
B.13	Bewegungsunschärfe bei EM-Aufnahmen	35
B.14	ctffind3 Diagnosebilder	36
B.15	Manuelle Selektion der Partikel	37
B.16	Kontrast Transfer Funktion korrigieren	39
B.17	Referenzfreie Ausrichtung	40
B.18	Verbessern der Signalqualität durch Summenbilder	41
B.19	Klassensummen von Proteinen	42
B.20	Euler Raumwinkel	43
B.21	<i>Sharpening</i> der 3D-Rekonstruktion mit Sparx	45
B.22	Anwenderfunktionsdatei für Sparx	46
B.23	Aus einem 3D-Modell erzeugte Referenzen	47
B.24	Fourier-Schalen-Korrelation (FSC)	49

ABBILDUNGSVERZEICHNIS

C.1	Vergleich des Hämoglobins von <i>Daphnia pulex</i> und <i>Daphnia magna</i>	56
C.2	Auflösungsbestimmung der 3D-Rekonstruktion von <i>D. pulex</i>	57
C.3	3D-Rekonstruktion des <i>D. pulex</i> -Hämoglobins mit 16 Untereinheiten	58
C.4	3D-Rekonstruktion des <i>D. pulex</i> -Hämoglobins mit 12 Untereinheiten	59
C.5	Kryo-EM-Aufnahme von <i>Astacus leptodactylus</i> -Hämocyanin, und ein Teil der Klassensummen	60
C.6	Auflösungsbestimmung der 3D-Rekonstruktion von <i>A. leptodactylus</i>	61
C.7	3D-Rekonstruktion des 2x6meren Hämocyanins aus <i>A. leptodactylus</i>	62
C.8	Atomares Modell des <i>A. leptodactylus</i> -Hämocyanins mit Symmetrieachsen	63
C.9	Kontaktstelle zwischen den beiden Hexameren des <i>A. leptodactylus</i> -Hämocyanins	63
C.10	Klassensummen des <i>Limulus polyphemus</i> -Hämocyanins	65
C.11	Auflösungsbestimmung der 3D-Rekonstruktion von <i>L. polyphemus</i>	66
C.12	3D-Rekonstruktion des <i>L. polyphemus</i> -Hämocyanins	67
C.13	Vermessung der 3D-Rekonstruktion des <i>L. polyphemus</i> -Hämocyanins	69
C.14	Flexibles <i>Fitting</i> mit <i>Direx</i>	71
C.15	IIIA–IIIA-Bindestelle der 7,1 Å-Rekonstruktion des <i>L. polyphemus</i> -Hämocyanins	73
C.16	IIIA–IIIA-Bindestelle der 7,4 Å-Rekonstruktion des <i>L. polyphemus</i> -Hämocyanins	74
C.17	IIIA–IIIA-Bindestelle der 10 Å-Rekonstruktion mit C1-Symmetrie des <i>L. polyphemus</i> -Hämocyanins	75
C.18	Position der IIIA–IIIA-Bindestelle im <i>L. polyphemus</i> -Hämocyanin	77
C.19	Fixierte und flexible Atome während der Molekulardynamik- Simulation der IIIA–IIIA-Bindestelle von <i>L. polyphemus</i>	78
C.20	Atome zur Vermessung in der MD-Simulation	80
C.21	Von <i>Chimera</i> erstellte Atomgruppen zur Verwendung in <i>Gromacs</i> .	80
C.22	Analyse der Molekulardynamik-Simulation	81
C.23	Darstellung des Balkendiagramms als Scheibendiagramm	82
C.24	Ergebnisse der MD-Simulation der IIIA–IIIA-Bindestelle	84
C.25	Ergebnisse der MD-Simulation der Intra-Hexamer-Bindestellen . .	85
C.26	Beteiligte Aminosäuren der IIIA–IIIA Bindestelle in Puffer A . . .	86
C.27	Beteiligte Aminosäuren der IIIA–IIIA Bindestelle in Puffer B . . .	88
C.28	Dichtekarte der IIIA–IIIA-Bindestelle, berechnet aus der MD-Simulation und auf 7 Å gefiltert	89
C.29	Dichtekarte der IIIA–IIIA-Bindestelle, berechnet aus der MD-Simulation und auf 10 Å gefiltert	90
C.30	Grafische Benutzeroberfläche - Hauptfenster	92
C.31	Konfigurationsdatei für <i>MDEMTols</i>	93
C.32	Grafische Benutzeroberfläche - Kategorien	94
C.33	DokuWiki auf dem Rechen-Cluster der Arbeitsgruppe	98

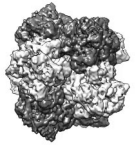


C.34	Programmcode zur Erstellung der Atomgruppen für Gromacs innerhalb von Chimera	100
C.35	Analyse der Partikelorientierung mit <code>anglespread2.py</code>	102
D.1	Homologiemodell einer Untereinheit des <i>Daphnia pulex</i> -Hämoglobins	107
D.2	<i>D. pulex</i> -Hämoglobin Sequenz einer Untereinheit	108
D.3	Vergleich von <i>Astacus leptodactylus</i> -Hämocyanin und 2x6mer aus <i>Limulus polyphemus</i> -Hämocyanin	111
D.4	Verschiebung in der IIIA–IIIA-Bindestelle unter Einfluss von $MgCl_2$	114
D.5	Fixierung der Aminosäuren für zukünftige MD-Simulationen	115
D.6	3D-Rekonstruktion des <i>L. polyphemus</i> -Hämocyanins ohne Symmetrie	117
D.7	<i>L. polyphemus</i> -Hämocyanin Vergleich von Martin et al. (2007) und diese Arbeit	119
D.8	Vergleich <i>L. polyphemus</i> -Hämocyanin der 7,1 Å-Rekonstruktion mit kristallographischem Hexamer	120
D.9	Vergleich <i>L. polyphemus</i> -Hämocyanin der 12 Å-Rekonstruktion mit kristallographischem Hexamer	121
D.10	Ausgabe von <code>readTifHead.py</code>	125
D.11	Lesbarkeit der Graphen von <code>gromacsgraph.py</code>	126
D.12	Fixieren ausgewählter Atome mithilfe von Chimera	127
D.13	FSC erstellt mit <code>plotfsc.py</code>	129
E.1	Aminosäurenabstände der 1. MD-Simulation im Puffer A	136
E.2	Scheibendiagramm aus Abb E.1	137
E.3	Aminosäurenabstände der 2. MD-Simulation im Puffer A	138
E.4	Scheibendiagramm aus Abb E.3	139
E.5	Aminosäurenabstände der 3. MD-Simulation im Puffer A	140
E.6	Scheibendiagramm aus Abb E.5	141
E.7	Aminosäurenabstände der 4. MD-Simulation im Puffer A	142
E.8	Scheibendiagramm aus Abb E.7	143
E.9	Aminosäurenabstände der 5. MD-Simulation im Puffer A	144
E.10	Scheibendiagramm aus Abb E.9	145
E.11	Aminosäurenabstände der 6. MD-Simulation im Puffer A	146
E.12	Scheibendiagramm aus Abb E.11	147
E.13	Aminosäurenabstände der 1. MD-Simulation im Puffer B	148
E.14	Scheibendiagramm aus Abb E.13	149



Tabellenverzeichnis

B.1	Parameter und Standorte der verwendeten Elektronenmikroskope	22
B.2	Verwendete freie Software	27
B.3	Verwendete kommerzielle Software	27
B.4	Verwendete Programmmodule und ihre Funktion	52
C.1	EM-Datensätze des <i>Daphnia pulex</i> -Hämoglobins	56
C.2	Kryo-EM-Datensatz des <i>Astacus leptodactylus</i> -Hämocyanins	61
C.3	Kryo-EM-Datensatz des <i>Limulus polyphemus</i> -Hämocyanins	64
C.4	Strukturelle Parameter des <i>L. polyphemus</i> -Hämocyanins	69
C.5	Direrx-Parameter für das flexible <i>Fitting</i>	70
C.6	Markup Befehle für den DokuWiki	97
D.1	Massenbestimmung bei <i>Daphnia</i> -Hämoglobinen	105
D.2	Abmessung unterschiedlicher 2x6mere	110
D.3	Vergleich der Sequenzen der Hämocyanine von <i>Panulirus interruptus</i> und <i>L. polyphemus</i> am IIIA–IIIA-Interface.	116
D.4	Abmessungen der <i>L. polyphemus</i> -Hämocyanin 3D-Rekonstruktionen	118
D.5	Abmessungen des <i>L. polyphemus</i> -Hämocyanins	118



A Einleitung

Proteinmoleküle sind etwa eine Million Mal kleiner als der Punkt am Ende dieses Satzes. Dennoch kann man sie im Transmissions-Elektronenmikroskop (EM) sehen und heute aus EM-Bildern ihre komplizierte dreidimensionale (3D) Struktur rekonstruieren. Dazu stellt die Bioinformatik spezielle Softwarepakete und Algorithmen bereit, die ständig weiterentwickelt werden (van Heel and Keegstra, 1981; van Heel et al., 1996; Tang et al., 2007; Ludtke et al., 1999; Hohn et al., 2007; Scheres, 2013, 2012).

Diese Arbeit beschreibt den Weg, auf dem drei extrazelluläre respiratorische Proteine mithilfe 3D-Elektronenmikroskopie (3D-EM) analysiert wurden. Dabei wurden unterschiedliche Auflösungen erreicht, je nachdem, welche Fragestellung zugrunde lag. In einem Fall wurden die 3D-EM-Ergebnisse durch Molekulardynamik-Simulationen (MD-Simulation) ergänzt. Um die Verknüpfung der eingesetzten Methoden zu vereinfachen, wurden von mir spezielle Software-Tools entwickelt und mittels einer leicht zu erweiternden grafischen Benutzeroberfläche (*graphical user interface*) (GUI) dem Anwender zur Verfügung gestellt.

1 Das Elektronenmikroskop als Werkzeug zur Proteinanalyse

Dank der geringen Wellenlänge der Elektronen, ist mit dem von Knoll and Ruska (1932) erfundenen und seither permanent weiterentwickelten EM (Abbildung (Abb.) A.1) im Prinzip eine atomare Auflösung möglich (Cosslett et al., 1979). Aufgrund der Elektronenempfindlichkeit biologischer Proben mussten jedoch spezielle Methoden entwickelt werden, damit Proteine im EM überhaupt untersucht werden können.

Eine schnelle und effiziente Methode ist die Negativkontrastierung von Pro-

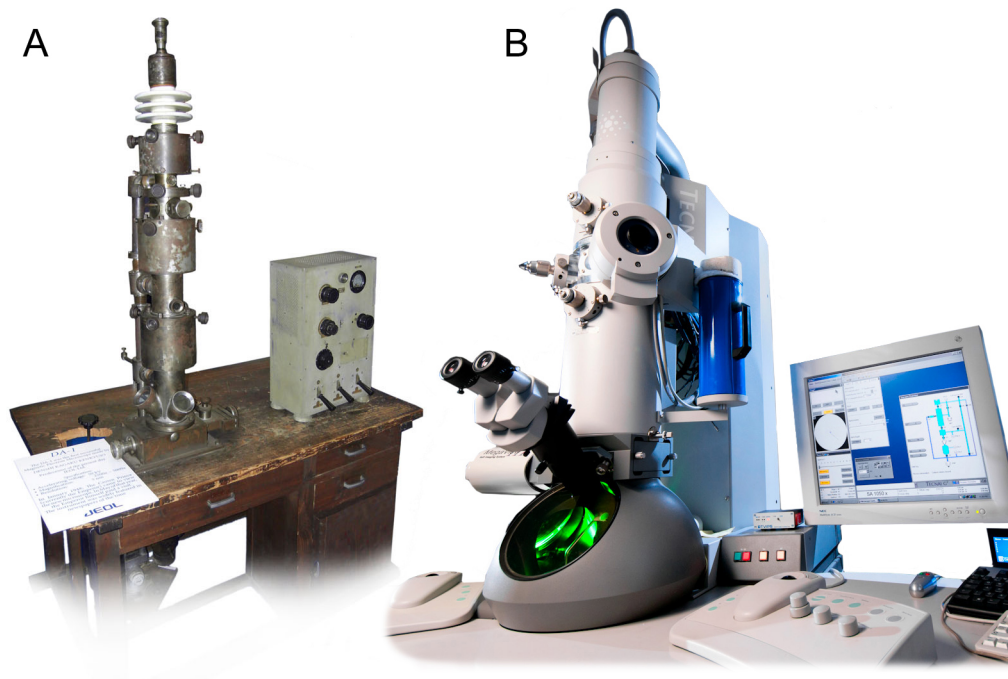


Abb. A.1: (A) DA-1 Transmissions-Elektronenmikroskop aus dem Jahr 1947 von der Firma JEOL (Japan). Es besaß eine Beschleunigungsspannung von 50 kV, arbeitete mit Vergrößerungen zwischen 2 500 und 5 000-fach und erreichte eine Auflösung von 5 nm. (B) Tecnai G² 12 BioTWIN TEM von der Firma FEI (Eindhoven, Niederlande) (Bild © Christoph Kühne). Es besitzt eine Beschleunigungsspannung von 20 bis 120 kV, arbeitet mit Vergrößerungen zwischen 22 und 300 000-fach und erreicht eine Auflösung von 0,34 nm.

teinen mit Schwermetallen (Brenner and Horne, 1959; Harris and Horne, 1994; Harris and Agutter, 1970; Horne, 1991). Die Proteinpartikel werden dabei mit Schwermetall überzogen, das sie vor der schädlichen Strahlung schützt. Ein Nachteil dabei ist, dass die Proteinmoleküle von den Schwermetallen verformt werden können und gleichzeitig die Auflösung durch den Schermetallmantel limitiert ist. Die von (Dubochet et al., 1988; Adrian et al., 1984) entwickelte Kryo-EM-Methode umgeht diese Schwierigkeiten. Dabei werden die Proteine schockartig in dünnes glasartiges Eis eingefroren (vitrifiziert). Sie behalten dadurch ihre natürliche Struktur und eine höhere Auflösung kann erreicht werden. Aufgrund der Strahlungsempfindlichkeit der Proteine sind allerdings nur sehr geringe Dosen an Elektronen möglich, damit die Proteine nicht beschädigt werden (Penczek et al.,

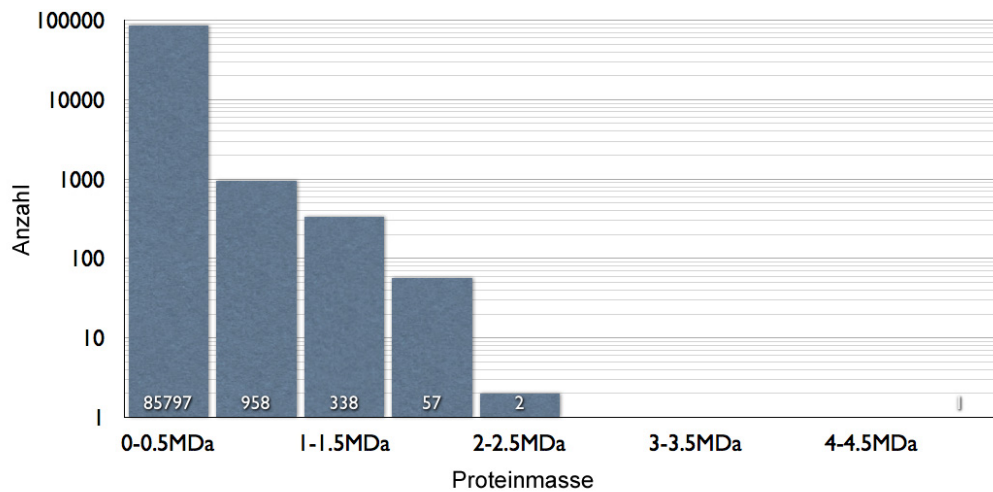
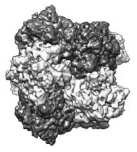


Abb. A.2: Die Anzahl atomarer Strukturen in der Protein Data Bank (PDB) (Berman et al., 2000), gruppiert nach kDa (Stand Januar 2013) (www.pdb.org, 2013). Die y -Achse ist logarithmisch. Etwa 98 % der publizierten Kristallstrukturen haben weniger als 500 kDa. Zwar sind nur wenige Proteine größer als 500 kDa, aber ihr weitgehendes Fehlen in der Datenbank spiegelt die Schwierigkeiten bei der Kristallisation wider.

1992; Dubochet et al., 1988). Außerdem sind die eingefrorenen Proteine äußerst kontrastarm (Zernike, 1955).

Um aus den zweidimensionalen (2D) EM-Aufnahmen eine gut aufgelöste 3D-Struktur des Proteins zu berechnen, müssen sie einer aufwändigen Bildverarbeitung unterzogen werden (Penczek, 2010a).

Die gängigste Methode zur Aufklärung der 3D-Struktur eines Proteins in atomarer Auflösung ist die Röntgenkristallographie (Kendrew and Parrish, 1957; Woolfson, 1970). Viele große und heterogene Proteinkomplexe konnten jedoch bisher nicht kristallisiert werden. Die Kristallisation von Molekülkomplexen mit mehr als 1000 kDa gelang bisher nur in wenigen Fällen (Abb. A.2). Gerade solche relativ großen Proteine sind jedoch für die Elektronenmikroskopie besonders geeignet. Ab einer Molekularmasse von etwa 400 kDa lassen sie sich mit 3D-EM gut bearbeiten (Henderson, 1995; Frank, 2002). Die notwendigen Schritte der Elektronenmikroskopie und der Bildverarbeitung bis zur fertigen 3D-Rekonstruktion sind in dieser Arbeit ausführlich beschrieben.

2 Respiratorische Proteine

Die meisten Lebewesen benötigen für ihren Energiestoffwechsel Sauerstoff. Während einfach gebaute Lebewesen genügend Sauerstoff über die Körperoberfläche aufnehmen können, brauchen komplexere Lebewesen ein inneres Sauerstofftransportsystem. Im Tierreich haben sich mehrere Proteine entwickelt, die den Sauerstoff von Haut, Lunge oder Kieme zu den Stellen im Körper transportieren, die ihn benötigen. Hier sind vor allem Hämoglobin und Hämocyanin zu nennen. Sie unterscheiden sich deutlich in ihrem aktiven Zentrum (Abb. A.3).

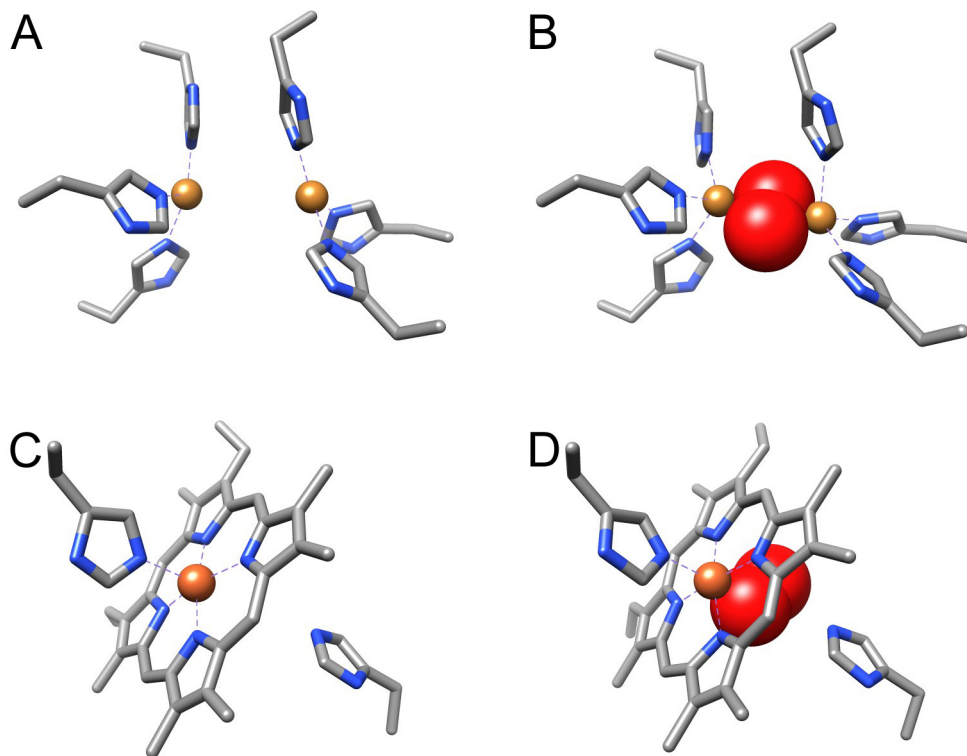
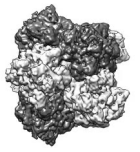


Abb. A.3: Vergleich des aktiven Zentrums von Hämoglobin und Hämocyanin. (A) Aktives Zentrum des Hämocyanins im deoxygenierten Zustand. Die beiden Kupferatome (gelb) werden durch die sechs Histidine koordiniert. (B) Aktives Zentrum des Hämocyanins mit gebundenem Sauerstoff (rot). (C) Aktives Zentrum des Hämoglobins im deoxygenierten Zustand. Das Fe^{2+} (orange) stellt das zentrale Ion in einem Protoporphyrinring sowie einem proximalen und einem distalen Histidin dar. (D) Das aktive Zentrum des Hämoglobins mit gebundenem Sauerstoff (rot).



Hämoglobin kommt in fast allen Tierstämmen vor. Es ist das Atmungsprotein der Wirbeltiere (Vertebraten), tritt aber auch in einzelnen Taxa der Platt-, Ringel-, Igel-, Hufeisen-, Bart- und Fadenwürmer auf und ebenso bei manchen Stachelhäutern, niederen Chordatieren, Weichtieren und Gliederfüßern (Weber and Vinogradov, 2001). Dabei gibt es große Unterschiede im Aufbau und der Struktur des Proteins. Einige Hämoglobine liegen intrazellulär vor, wie z. B. bei Wirbeltieren und Zuckmücken (Chironomidae) (Weber and Vinogradov, 2001). Diese Hämoglobine sind sehr klein, mit nur zwei oder vier Häm-Domänen. Oft sind aber auch extrazelluläre Hämoglobine vorhanden, die dann sehr groß sind. Das am Besten untersuchte Beispiel ist das 3600 kDa Hämoglobin des Regenwurms *Lumbricus terrestris* mit 144 Häm-Domänen (Mouche et al., 2001; Schatz et al., 1995).

Hämoglobine sind alle aus einem Grundbaustein aufgebaut, einer Globinkette mit einer Häm-Gruppe. Hier kann es durch Genduplikationen dazu kommen, dass nicht nur eine Häm-Domäne pro Polypeptidkette vorhanden ist, sondern auch zwei oder mehrere (Weber and Vinogradov, 2001). Die einzelne Häm-domäne besitzt in der Regel acht α -Helices und besteht aus etwa 150 Aminosäuren; die Molekularmasse beträgt ca. 17 kDa (Abb. A.4).

Im Gegensatz zu den Hämoglobinen, in den sich die Quartärstruktur zwischen verschiedenen Taxa deutlich unterscheidet, gibt es bei durchweg extrazellulären Hämocyaninen nur zwei Grundtypen. Zum einen existiert das zylinderförmige Molluskenhämocyanin mit einem Durchmesser von 35 nm (Markl, 2013). Zum anderen gibt es das wahrscheinlich konvergent entstandene Arthropoden-Hämocyanin, welches sich aus einem hexameren Grundbaustein aufbaut (Markl and Decker, 1992). Pro Untereinheit ist ein aktives Zentrum beim Arthropoden-Hämocyanin vorhanden. Im Gegensatz zum roten Hämoglobin ist Hämocyanin im oxygenierten Zustand blau und im desoxygenierten Zustand farblos. Neben Hexameren (1x6meren) sind bei Arthropoden 2x6mere, 4x6mere, 6x6mere und 8x6mere bekannt. Den derzeitigen Stand der Quartärstruktur-Forschung geben folgende Publikationen wieder: 1x6mer: (Volbeda and Hol, 1989); 2x6mer: (Stöcker et al., 1988; Mičetić et al., 2010); 4x6mer: (Markl et al., 1981; Cong et al., 2009); 6x6mer: (Markl et al., 2009); 8x6mer: (Martin et al., 2007).

Wie Hämoglobin zeigt auch Hämocyanin eine deutliche Kooperativität (Markl and Decker, 1992). Das bedeutet, wenn eine Untereinheit einen Liganden (in der

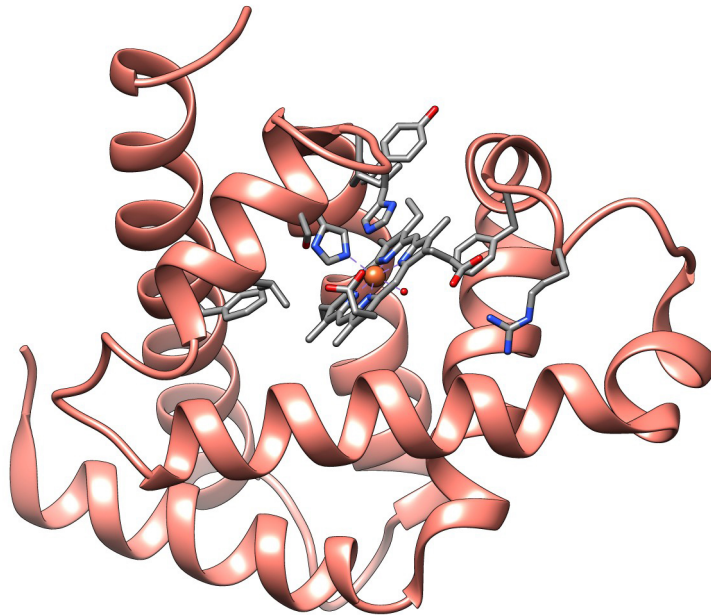


Abb. A.4: Struktur einer Untereinheit bzw. einer Häm-Domäne (PDBid 1MBN).

Regel Sauerstoff) bindet, dann steigt die Affinität der übrigen Untereinheiten, ebenfalls ein Sauerstoffmolekül zu binden. Das funktioniert ebenso in entgegengesetzte Richtung: Verliert ein voll beladenes Hämocyanin ein Sauerstoffmolekül, gibt es die restlichen noch gebundenen O_2 -Moleküle leichter wieder ab.

2.1 Das Hämoglobin des Wasserfloh *Daphnia pulex*

Das Hämoglobin des Wasserfloh *D. pulex* ist bis auf die Quartärstruktur gut untersucht. In sauerstoffarmer Umgebung steigern die Tiere die Expression des Hämoglobins um einen Faktor von 15 bis 20 (Fox, 1948; Zeis et al., 2003; Colbourne et al., 2011). Im Gegensatz zu Säugetieren erhöhen sie die O_2 -Affinität über die Expression anderer Peptidketten und nicht über die Bindung von Liganden (Gerke et al., 2011). Die Tiere färben sich dann blutrot. Dieses Schauspiel ist in der freien Natur zu sehen, wenn die Teiche und Tümpel, in denen die Tiere leben, durch die Sonne erwärmt werden und der Sauerstoffgehalt dadurch sinkt. Der Hill Koeffizient von *D. pulex* liegt zwischen 1.4 und 2.4 (Wolf et al., 1983).

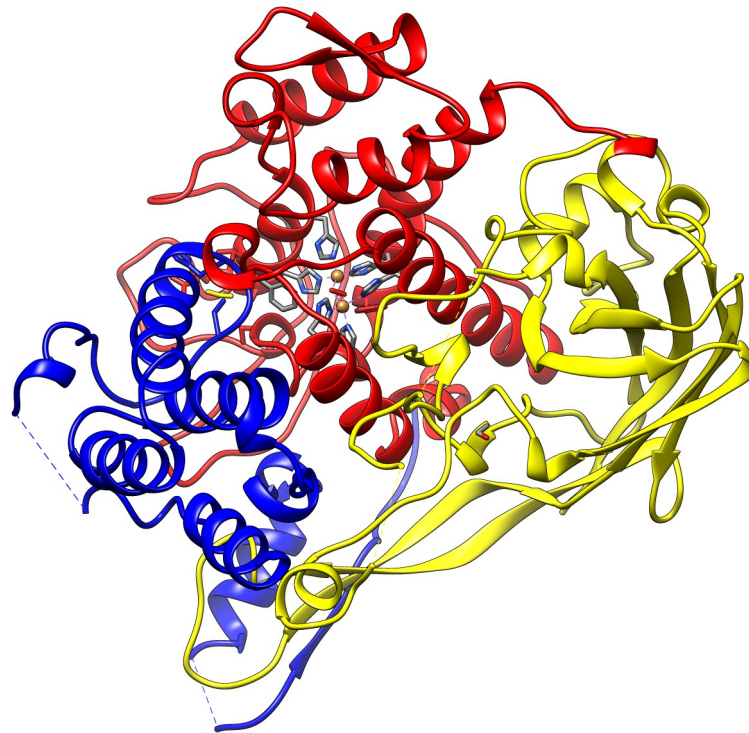
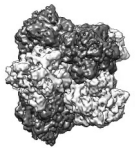


Abb. A.5: Struktur einer Untereinheit des Artropoden-Hämocyanins (PDBid 1NOL). Domäne 1 (blau), Domäne 2 (rot) und Domäne 3 (gelb).

Die Quartärstruktur des *Daphnia*-Hämoglobins ist weitgehend unklar. Die Polypeptidkette (ca. 36 kDa) umfasst zwei Hämdomänen. Für die Quartärstruktur wurden 12 oder 14 solcher Untereinheiten vorhergesagt (Dewilde et al., 1999; Peeters et al., 1990).

Parallel wurde beim Wasserfloh *Daphnia magna* ebenfalls das Hämoglobin untersucht und eine deutlich höhere Masse festgestellt (Sugano and Hoshi, 1971; Ilan et al., 1982; Lamkemeyer et al., 2005). Dabei ergaben sich abhängig von der verwendeten Analyseverfahren zur Proteinmassenbestimmung große Unterschiede. Eine 3D-Rekonstruktion von *D. magna*-Hämoglobin mit einer Auflösung von ca. 30 Å wurde publiziert (Lamkemeyer et al., 2006), wick aber von dem bei *D. pulex* erstellten Modell deutlich ab. Mittels 3D-EM sollte ich hier Klarheit schaffen.

2.2 Das Hämocyanin des Flusskrebse *Astacus leptodactylus*

Das Hämocyanin von *A. leptodactylus* ist aus zwei Hexameren (1x6meren) aufgebaut (Markl et al., 1979; Stöcker et al., 1988). Die Quartärstruktur des 1x6mers ist im Detail bekannt (Abb. A.6) (Volbeda and Hol, 1989). Die Anordnung der 1x6mere zueinander war jedoch unbekannt. Hier sollte eine Disulfidbrücke eine Rolle spielen (Markl et al., 1981). Meine Aufgabe war zu ermitteln, wie die beiden 1x6mere assoziiert sind.

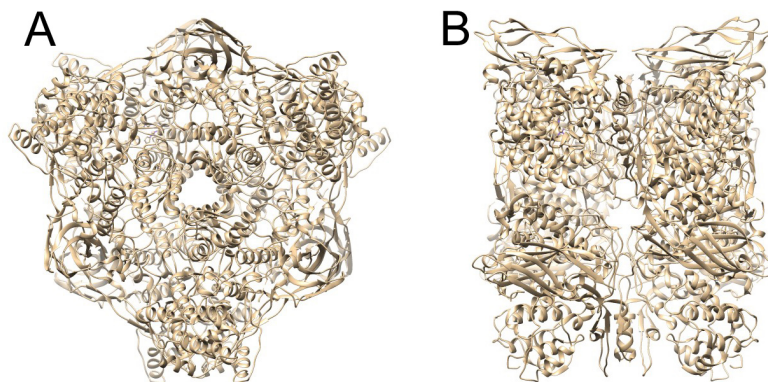


Abb. A.6: Struktur des 1x6meren Hämocyanins von *Panulirus interruptus* (PDBid: 1HCY; (Volbeda and Hol, 1989)).

2.3 Das Hämocyanin des Pfeilschwanzes *Limulus polyphemus*

Bei dem Hämocyanin von *L. polyphemus* handelt es sich um das größte bekannte Arthropoden-Hämocyanin. Hier ist die Quartärstruktur mit einer Auflösung von 10 Å bekannt (Abb. A.7), und es existiert ein pseudoatomares Modell (Martin et al., 2007), was nach heutigen Kriterien etwa 12 Å entspricht. Es sollte eine verbesserte Rekonstruktion erstellt werden, um die Kontaktstellen zwischen den acht Hexameren genauer zu identifizieren.

Dadurch sollte der Weg bereitet werden, die chemomechanischen Mechanismen der Kooperativität dieses Hämocyanins genauer zu analysieren. Die Kooperativität ist bei diesem Hämocyanin mit einem Hill-Koeffizienten von 9 besonders hoch.

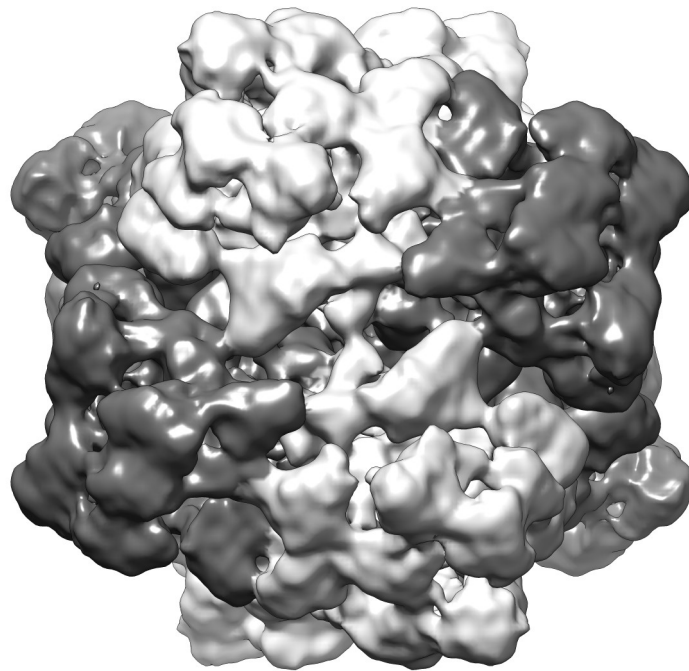
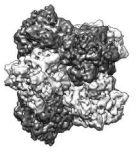


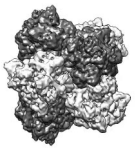
Abb. A.7: Struktur des *Limulus polyphemus*-Hämocyanins aus Martin et al. (2007). Kennnummer: 1304 in EMDDataBank (Lawson et al., 2011).

3 Bioinformatische Methoden in der 3D-EM

In der 3D-Elektronenmikroskopie müssen große Datensätze von über 10 000 Bildern verarbeitet werden. Den verfügbaren Programmen fehlt häufig eine grafische Benutzeroberfläche (*graphical user interface*) (GUI). Sie werden über eine Kommandozeile bedient und sind daher nicht sehr intuitiv. Hier ist es von großem Interesse eine leicht erweiterbare GUI zu erstellen, die den Benutzer in seiner Arbeit unterstützt und einfach zu bedienen ist. Zudem werden mehrere alternative Programmpakete benutzt, die ähnliches leisten, aber jeweils Vor- und Nachteile haben.

Um nur die Vorteile der verschiedenen Programme nutzen zu können, ist oft eine aufwendige Neuformatierung der Daten nötig, damit sie von mehreren Programmen gleichzeitig genutzt werden können. Dabei werden häufig bioinformatische Werkzeuge benötigt, die bei der Formatierung und Auswertung der Daten behilflich sind, die aber noch nicht existieren.

Ziel dieses Teils meiner Arbeit war es, durch Entwicklung solcher kleinen Hilfsprogramme die *Workflows* der Arbeitsgruppe zu optimieren. Die ebenfalls zu entwickelnde GUI sollte dann ein Paket bilden, das alle diese Tools zur Verfügung stellt.

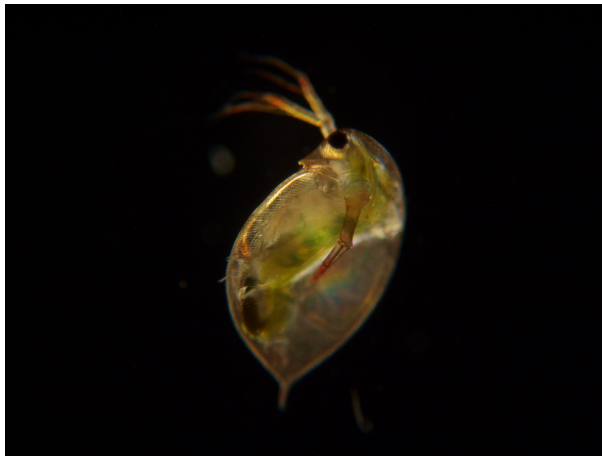


B Material und Methoden

1 Tiere

In dieser Arbeit wurden die respiratorischen Proteine verschiedener Arthropoden strukturell untersucht. Die Proteine liegen alle extrazellulär in einem offenen Blutkreislaufsystem vor. Dabei ist *Daphnia pulex*, der Gemeine Wasserfloh, einer der wenigen Gliederfüßer, die als Atmungsprotein Hämoglobin anstatt Hämocyanin besitzt. Die Tiere sowie ihr Lebensraum werden im Folgenden kurz vorgestellt.

Daphnia pulex



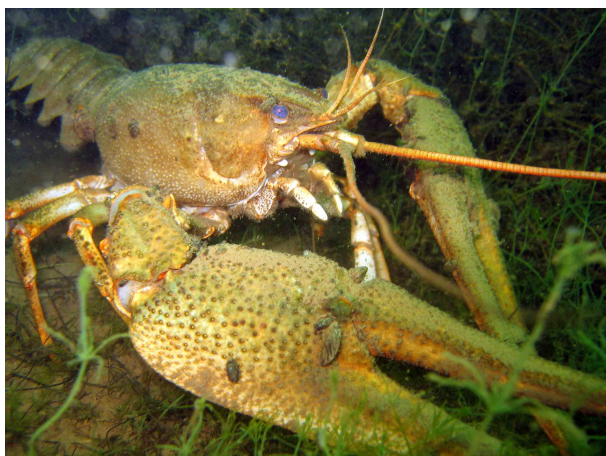
Stamm:	Arthropoda
Unterstamm:	Crustacea
Klasse:	Branchiopoda
Ordnung:	Onychora
Familie:	Daphniidae
Gattung:	<i>Daphnia</i>
Art:	<i>pulex</i>

Abb. B.1: Der Gemeine Wasserfloh *Daphnia pulex* und das zugehörige System (Bild © Yann Reydelet, Uni Münster)

Der Gemeine Wasserfloh *Daphnia pulex* (Leydig, 1860) besitzt eine Körperlänge von ca 3 bis 4 mm, wobei die Männchen unter 2 mm bleiben (Abb. B.1). Er lebt oft in kleinen Tümpeln oder an Randregionen großer Seen. Im Unterschied

zu vielen anderen Crustacea besitzt *D. pulex* eisenhaltiges Hämoglobin anstelle von kupferhaltigem Hämocyanin. Sie können, abhängig vom Sauerstoffgehalt im umgebenden Medium, die Hämoglobinkonzentration im Körper regulieren (Gerke et al., 2011). Bei geringer Sauerstoffkonzentration färben sich die Tiere tief rot aufgrund der erhöhten Expression von Hämoglobin. Unter normoxischen Bedingungen sind sie weiß bis transparent. Dieses Phänomen kann auch in natürlichen Lebensräumen beobachtet werden, da sich die oberen Wasserschichten im Sommer stark aufwärmen und dadurch der Sauerstoffgehalt abnimmt.

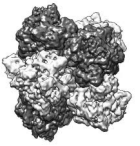
Astacus leptodactylus



Stamm:	Arthropoda
Unterstamm:	Crustacea
Klasse:	Malacostraca
Ordnung:	Decapoda
Familie:	Astacidae
Gattung:	<i>Astacus</i>
Art:	<i>leptodactylus</i>

Abb. B.2: Der Galizische Sumpfkrebs *Astacus leptodactylus* und das zugehörige System (Bild © wikipedia: Ullrich Mühlhoff)

Der Galizische Sumpfkrebs *Astacus leptodactylus* (Eschscholz, 1823) erreicht ein Körpergröße von bis zu 18 cm und ist meist gelb bis braun gefärbt (Abb. B.2). Manche Tiere können auch eine grünliche Färbung aufweisen. Er bevorzugt hauptsächlich warme, langsam fließende oder stehende Gewässer, die sehr nährstoffreich sind. Für den Sauerstofftransport besitzen die Tiere ein 2x6meres Hämocyanin, das abhängig vom Häutungszustand der Tiere auch als 1x6mer vorliegen kann (Spindler et al., 1992).



Limulus polyphemus



Stamm:	Arthropoda
Unterstamm:	Chelicerata
Klasse:	Merostomata
Ordnung:	Xiphosura
Familie:	Limulidea
Gattung:	<i>Limulus</i>
Art:	<i>polyphemus</i>

Abb. B.3: Der Atlantische Pfeilschwanzkrebs *Limulus polyphemus* und das zugehörige System (Bild © wikipedia: Didier Descouens)

Der Atlantische Pfeilschwanzkrebs *Limulus polyphemus* (Linnaeus, 1758) besitzt eine Körperlänge von bis zu 60 cm (inkl. Schwanz), Männchen bleiben ca. 30 % kleiner (Abb. B.3). Die Verbreitung erstreckt sich hauptsächlich über den Golf von Mexiko und entlang der Atlantikküste von Nordamerika. Die Tiere leben dort an flachen Sandküsten und durchsuchen den Boden nach Schnecken und Muscheln. *L. polyphemus* ist mit den Spinnentieren und nicht mit den Krebstieren verwandt, wie fälschlicherweise durch den Namen angenommen werden könnte. Die Tiere besitzen ein 8x6meres Hämocyanin, das für den Sauerstofftransport verantwortlich ist (Brenowitz et al., 1981; Lamy et al., 1983; Bijlholt et al., 1979).

2 Biochemische Methoden

2.1 Isolierung der Proteine

Daphnia pulex

Die Aufreinigung des Hämoglobins von *D. pulex* wurde von [REDACTED] [REDACTED] [REDACTED] [REDACTED] [REDACTED] vorgenommen. Hierfür wurde ca. 1 g Tierextrakt homogenisiert. Phenylmethylsulfonylfluorid (PMSF) wurde in einer finalen Konzentration von 1 mM zugegeben, um einen proteolytischen Abbau zu verhindern. Der flüssige Teil des Extraktes wurde bei 12 000 g für 15 Minuten

zentrifugiert (Zeis et al., 2003; Lamkemeyer et al., 2006). Das Hämoglobin fand sich im Überstand und wurde entweder sofort verwendet oder maximal 5 Tage bei -20°C gelagert. Zur Verdünnung wurde ein 50 mM TRIS Puffer mit pH 8 benutzt (siehe Kapitel E.4).

Astacus leptodactylus

Die Probenentnahme geschah ventral aus der Bauchlaktone. Hierbei wurde mit einer Spritze etwa 0,5 ml Hämolymphe entnommen indem die Kanüle flach zwischen Abdomen und Cephalothorax eingestochen wurde. Die Zellen wurden bei 12 000 g abzentrifugiert und der Überstand danach in einer präparativen Ultrazentrifuge für drei Stunden bei 130 000 g zentrifugiert. Das pelletierte Hämocyanin wurde mit Puffer (siehe Kapitel E.4) über Nacht resuspendiert.

Um die 1x6mere von den 2x6meren zu trennen, wurde eine Gelfiltrationschromatographie (siehe Kapitel B.2.2) verwendet. Diese liefert stark angereicherte 2x6mere, die für die Experimente benutzt wurden.

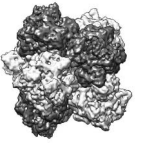
Limulus polyphemus

Die Entnahme von Hämolympfproben wurde von [REDACTED] [REDACTED] und [REDACTED] [REDACTED] durchgeführt. Hierfür wurde das Tier mit einer Kanüle an der dorsalen Gelenkhaut zwischen Schwanz und Carapax punktiert und ca 1 ml Hämolymphe entnommen. Diese wurde für 5 Minuten bei 12 000 g zentrifugiert um die Blutzellen zu sedimentieren. Der Überstand wurde in ein neues Gefäß mit ca 50 % Puffer (siehe Kapitel E.4) überführt und für die Versuche verwendet.

2.2 Gelfiltrationschromatographie

Das Hämocyanin von *A. leptodactylus* besteht aus einem 1x6mer oder 2x6mer, abhängig von verschiedenen Faktoren (Decker and Föll, 2000). Für diese Arbeit war hauptsächlich das Dihexamer von Interesse und sollte aus der Hämolymphe mithilfe des Größenausschlussverfahrens aufgereinigt werden.

Dazu wurde die Proteinlösung mit einem geeigneten Puffer auf eine Säule aufgetragen, die mit einem inerten polymeren Material gefüllt war. Dieses Material hat durch die Produktion eine bestimmte Porengröße. Große Proteine umfließen diese Poren, wohingegen kleinere Proteine in die Poren einwandern



und somit eine längere Strecke durch die Säule zurücklegen (Ackers, 1970). Die Trenngrenzen lassen sich durch das verwendete Material verändern. Durch eine längere Trennstrecke kann die Trennschärfe verbessert werden.

Verwendet wurde eine Glassäule mit 1 cm Innendurchmesser und 60 cm Länge. Das Filtermaterial war das Agarosegel Bio-Gel A-1.5m, Coarse von Bio-Rad (München, Deutschland).

3 Elektronenmikroskopie

Um die verschiedenen optischen Fehler besser zu verstehen, erläutert dieser Abschnitt einige grundlegende optische Phänomene, die bei der Verarbeitung elektronenmikroskopischer Bilder (EM-Bilder) von Proteinen korrigiert werden müssen.

3.1 Allgemeine Grundlagen

Ob man ein Objekt der Größe y sehen kann ist abhängig vom Betrachtungsabstand S und dem Sehwinkel σ und wird durch folgendes Verhältnis ausgedrückt:

$$\tan \sigma = \frac{y}{S} \text{ bzw. bei kleinen Winkeln } (\sigma \ll 1): \sigma = \frac{y}{S}$$

In der Physik geht man von einem idealen Betrachtungsabstand von $S = 25$ cm (deutliche Sehweite) aus (Thomas and Gemming, 2013). Wenn wir dieses Verhältnis auf Proteine anwenden, die eine Größe von wenigen Nanometern besitzen, dann entspricht das etwa dem Versuch eine Stubenfliege in 600 km Entfernung erkennen zu wollen. Mithilfe eines Mikroskops können wir den Sehwinkel durch Linsensysteme vergrößern ohne den Betrachtungsabstand zu verändern. Die Linsen lenken dabei die Photonen bzw. Elektronen in einem neuen Winkel ab (Abb. B.4). Dadurch erhalten wir eine Vergrößerung M , die wie folgt definiert ist:

$$M = \frac{\text{Sehwinkel mit optischem Gerät}}{\text{Sehwinkel ohne optisches Gerät}} = \frac{\sigma'}{\sigma}$$

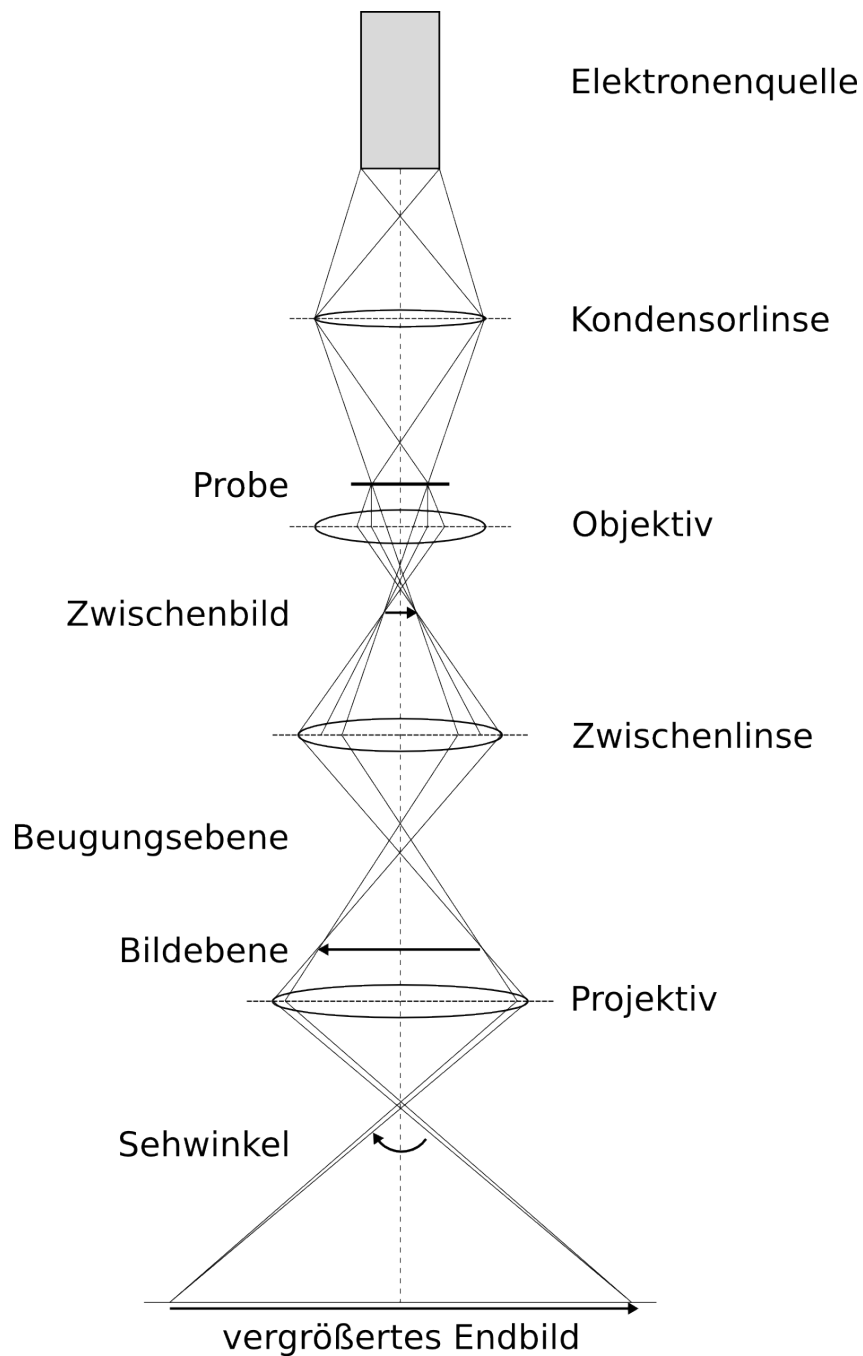
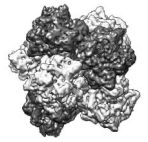


Abb. B.4: Prinzip eines Elektronenmikroskops. Durch elektromagnetische Linsen mit variabler Brennweite lässt sich der Sehwinkel fast beliebig verändern. In der Elektronenmikroskopie sind Primärvergrößerungen von bis zu 120 000-fach keine Seltenheit.



Optische Auflösung

Mit einem Mikroskop lässt sich ein Objekt nur bis zu einem bestimmten Faktor vergrößern. Die Grenzen werden vereinfacht durch das Rayleigh-Kriterium für sichtbares Licht beschrieben (Williams and Carter, 2009):

$$\delta = \frac{0,61 \cdot \lambda}{n \cdot \sin \alpha}$$

(δ : Abstand zweier noch sichtbaren Punkte (Punktauflösung), λ : Wellenlänge, n : Brechungsindex des umgebenden Mediums, $\sin \alpha$: Beugungswinkel der Welle) Damit erhalten wir für sichtbares Licht eine Punktauflösung von etwa $0,2 \mu\text{m}$.

Durch die Verwendung von Elektronen im Transmissionselektronenmikroskop (TEM) wird infolge der geringeren Wellenlänge eine höhere Auflösung erreicht. Die Wellenlänge der Elektronen ist abhängig von der Beschleunigungsspannung U_B . Für eine Beschleunigungsspannung über 50 kV sind relativistische Effekte aufgrund der Geschwindigkeit nicht mehr zu vernachlässigen und die Wellenlänge der Elektronen lässt sich wie folgt berechnen (Williams and Carter, 2009):

$$\lambda = \frac{h}{\sqrt{e \cdot U_B \left(2 \cdot m_0 + \frac{e \cdot U_B}{c^2}\right)}}$$

(h : Plancksches Wirkungsquantum, e : Elementarladung, m_0 : Ruhemasse, c : Lichtgeschwindigkeit im Vakuum)

Daraus errechnet sich für eine Beschleunigungsspannung von 300 kV eine Wellenlänge von $1,97 \text{ pm}$ und eine Auflösungsgrenze von etwa $0,016 \text{ \AA}$. Zur Erinnerung: Atome mit samt Elektronenhülle haben Durchmesser von 1 bis 5 \AA (Williams and Carter, 2009).

Technische Einschränkungen

Dieser Wert wird wegen mehrerer Faktoren in der Praxis nicht erreicht, dabei ist das größte Problem bei biologischen Proben die Schädigung durch den Elektronenstrahl. Realistische Werte für Metalle sind zur Zeit etwa $0,5 \text{ \AA}$ (Kisielowski et al., 2008), bei Proteinkomplexen liegt die Grenze bei 3 \AA bis 4 \AA (Yu et al., 2011; Zhang et al., 2010; Estrozi and Navaza, 2010; Baker et al., 2013). Wobei hier allerdings nicht die Punktauflösung gemeint ist, sondern Unterschiede

zwischen Datensätzen (siehe Kapitel B.6.10). Die Punktauflösung ist in der Regel deutlich geringer.

Die Einschränkungen kommen unter anderem aufgrund verschiedener Abbildungsfehler zustande. Da sich biologische Proben im TEM wie Phasenobjekte verhalten, das heißt (d. h.) die Probe keinen Intensitätsunterschied der Elektronen verursacht, sind diese im Fokus nicht zu erkennen (Zernike, 1953). Durch defokussieren erreicht man eine Phasenverschiebung zwischen den durch die Probe gebeugten und den ungebeugten Elektronen. Das führt dazu, dass Strukturen durch die Punktspreizfunktion (*point spread function*) (PSF) dargestellt werden und die Abbildungen nicht direkt interpretierbar sind. Abhängig vom abgebildeten Abstand in der Abbildung (Raumfrequenz q) entsteht ein unterschiedlicher Kontrast innerhalb der Abbildung (Abb. B.5). Der Kontrast wird näherungsweise durch die Kontrast-Transfer-Funktion (*contrast transfer function*) (CTF) beschrieben, die der fouriertransformierten PSF entspricht (Williams and Carter, 2009):

$$CTF(q) = \sin\left(\frac{\pi}{2} (C_s \cdot \lambda^3 \cdot q^4 - 2 \cdot \Delta f \cdot \lambda \cdot q^2)\right) \cdot e^{-\pi^2 \cdot C_c^2 \cdot \left(\frac{\Delta E}{E_0}\right)^2 \lambda^2 q^4}$$

(q : Raumfrequenz, C_s : Öffnungsfehlerkonstante, λ : Elektronenwellenlänge, Δf : Defokus, C_c : Farbfehlerkonstante, ΔE : Energiebreite der Elektronen, E_0 : Primärelektronenenergie)

In der digitalen Bildverarbeitung wird der Einfluss der CTF auf die Abbildung herausgerechnet. Dadurch entspricht die korrigierte Abbildung idealerweise der echten Probe (siehe Abschnitt B.6.2).

3.2 Ausrüstung

Objektträger

Für das EM kann der Glasobjektträger der Lichtmikroskopie nicht benutzt werden, da er für Elektronen undurchlässig ist. Hier werden spezielle Objektträger benötigt, die für Elektronen möglichst durchlässig sind. Etabliert hat sich hier ein dünner Kohlenstofffilm, der über ein Kupfernetz gespannt wird (Abb. B.6). Der Kohlenstofffilm ist für Elektronen fast kein Hindernis, das Kupfernetz dient lediglich der Stabilisierung.

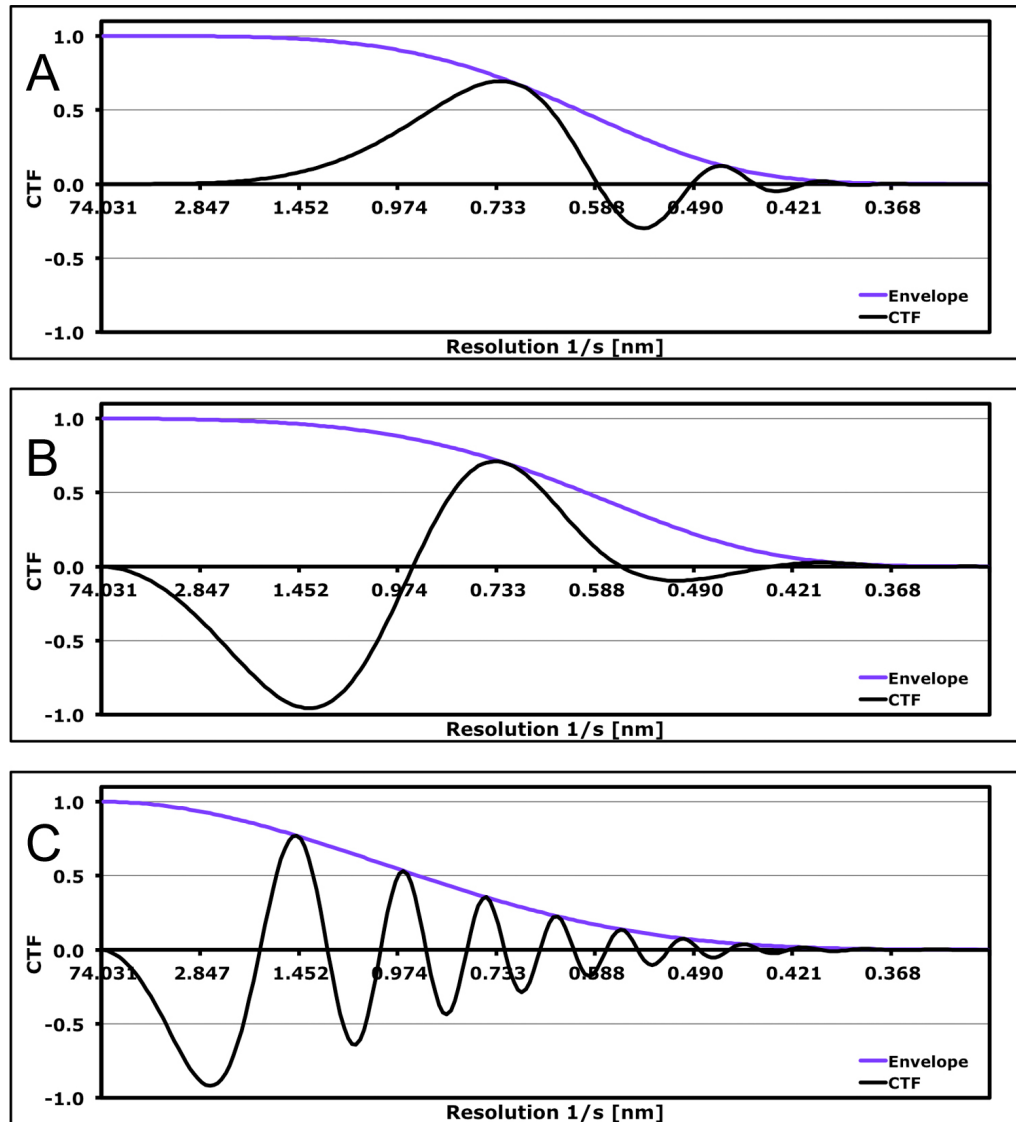
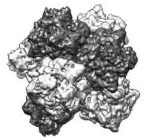


Abb. B.5: TEM Simulation der CTF mithilfe der Excel-Datei *CTF Calculation in Excel* (©Henning Stahlberg (Uni Basel, Schweiz)). Hier wird deutlich, wie der Kontrast (y -Achse) in Abhängigkeit der Raumfrequenz (x -Achse) moduliert. Durch die Defokussierung verändert sich der Kontrast der jeweiligen Raumfrequenz. (A) Im Fokus; (B) mit 300 nm Defokus; (C) mit 1000 nm Defokus.

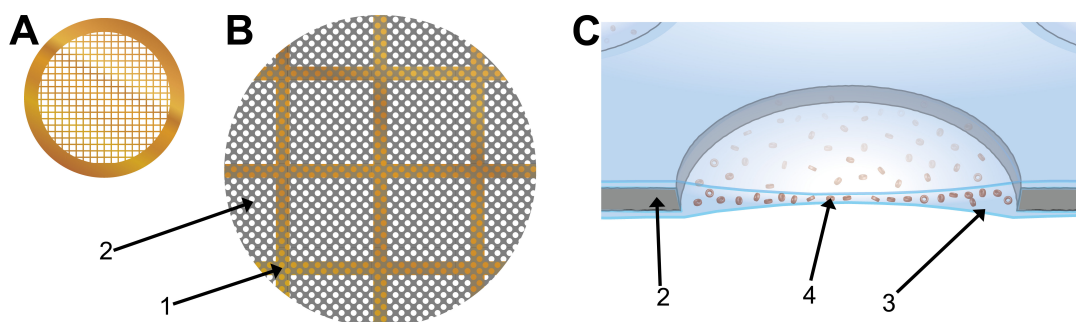


Abb. B.6: Schematische Darstellung eines Objektträgers mit löchrigem Kohlefilm. (A) Kompletter Objektträger (3 mm Durchmesser); (B) Ausschnitt mit Löchern; (C) Seitenansicht eines Lochs mit Proteinprobe in einer dünnen Eisschicht. (1) Kupfernetz; (2) löchriger Kohlefilm; (3) vitrifiziertes Eis; (4) Proteine. (Bild © Christoph Kühne)

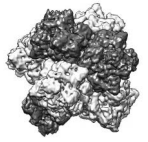
Abhängig von der Präparationsmethode (siehe Abschnitt B.3.3) werden unterschiedliche Anforderungen an die Objektträger gestellt. Bei der Negativkontrastierung wird ein durchgängiger Kohlefilm verwendet, an den die Proteine binden können. Für die Kryo-Präparation benutzt man häufig Kohlefilme mit Löchern, in denen die gelösten Proteinmoleküle in einer dünnen vitrifizierten Pufferschicht liegen. Das bietet den Vorteil, dass die Proteine ohne Hintergrund (Kohlefilm) vorliegen.

Glimmentladung (*Glow Discharge*)

Um die Bindung der Proteine an den Objektträger zu erleichtern, wird dieser meist hydrophiliert. Dies geschieht mithilfe der Glimmentladung (*glow discharge*). Die Glimmentladung wurde mit einem Emitech K100X Gerät der Firma Quorum Technologies (Guelph, Ontario, Kanada) durchgeführt. Die eingestellte Stromstärke betrug 25 mA und wirkte für 25 Sekunden.

Cryoplunger™3

Für die Kryo-Probenpräparation (Abschnitt B.3.3) wurde ein Cryoplunger™3 (CP3) der Firma Gatan (Pleasanton, Kalifornien, USA) verwendet. Dieser wurde um eine Atmosphärenkammer (entwickelt von [REDACTED]) erweitert, die es ermöglicht die Proteine vor der Präparation mit einem



spezifischen Gasgemisch zu equilibrieren. Diese Funktion ist besonders interessant für sauerstoffbindende Proteine wie Hämoglobin und Hämocyanin. Mithilfe dieser Kammer lässt sich die oxygenierte bzw. deoxygenierte Form des Proteins erzeugen und vitrifizieren.

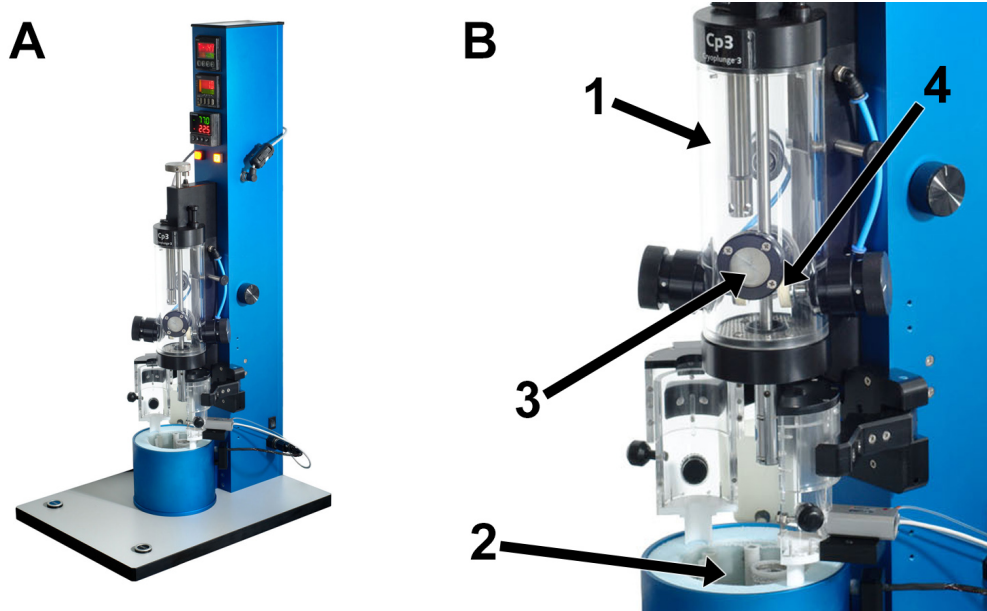


Abb. B.7: CryoplungerTM 3 mit Atmosphärenkammer zur Präparation von kryogenischen Proben. (A) Zeigt das Gerät in der Gesamtansicht; (B) fokussiert auf die Atmosphärenkammer (1) und auf den Behälter für flüssigen Stickstoff und flüssiges Ethan (2). Weiterhin sieht man die Probenöffnung (3) und die Stempel mit Filterpapier (4).
(Bild © Christoph Kühne)

Elektronenmikroskope

Die in dieser Arbeit verwendeten Elektronenmikroskope wurden alle von der Firma FEI (Eindhoven, Niederlande) hergestellt. Alle Geräte können sowohl mit einem Raumtemperaturprobenhalter für negativkontrastierte Proben, als auch mit einem speziellen Kryo-Probenhalter betrieben werden. Die in dieser Arbeit gezeigten negativkontrastierten Aufnahmen wurden alle an einem Tecnai G² 12 BioTWIN (Abb. B.8) im Institut für Zoologie der Johannes Gutenberg-Universität (JGU) Mainz vorgenommen. Eine Liste mit den mikroskopspezifischen Parametern zeigt Tabelle B.1.

Tab. B.1: Parameter und Standorte der verwendeten Elektronenmikroskope

	Tecnai G ² Polara	Tecnai G ² F20	Tecnai G ² 12 BioTWIN
Standort	MPI für Biophysik Frankfurt	MPI für Polymerforschung Mainz	Institut für Zoologie Mainz
Hersteller	FEI	FEI	FEI
Beschleunigungsspannung	300 kV	200 kV	120 kV
sphärische Aberration (C_s)	2,0 mm	2,0 mm	6,3 mm
chromatische Aberration (C_c)	2,2 mm	2,0 mm	5,0 mm
Detektor Typ	Kodak Film	Kodak Film	TVIPS F416
verwendet für	<i>L. polyphemus</i>	<i>D. pulex</i>	<i>A. leptodactylus</i>
verwendete Vergrößerung	61 400	62 950	71 540
gescannt bei	3256 DPI	4097 DPI	–
Pixelgröße	0,96 Å/Pixel	0,985 Å/Pixel	2,18 Å/Pixel

3.3 Probenherstellung für das Elektronenmikroskop

Um Proteine im EM zu analysieren, bedarf es spezifischer Präparationsmethoden. Abhängig vom gewünschten Ergebnis gibt es bei jeder Methode Vor- und Nachteile. Das generelle Ziel ist es aber, das Protein in einer möglichst natürlichen Form zu fixieren. Dabei sollten die Proteinmoleküle möglichst zahlreich sein und dennoch einzeln vorliegen, damit sie sich nicht gegenseitig verdecken.

Negativkontrastierung

Die häufigste Methode der Untersuchung von Proteinen im EM ist die Negativkontrastierung mithilfe von Schwermetallen (Brenner and Horne, 1959; Harris and Agutter, 1970). Hierbei legen sie sich wie ein Mantel um die Proteine. Die Schwermetalle sorgen für eine starke Ablenkung der Elektronen und somit für einen stärkeren Amplitudenkontrast der Probe. Die Proteinmoleküle sind dadurch besser zu erkennen, doch dies wirkt sich negativ auf die Struktur aus, die durch die Dichte der Schwermetallatome nicht höher aufgelöst werden kann (Harris and Horne, 1994). Weiterhin können durch das Austrocknen der Probe Artefakte entstehen, vor allem eine Abflachung (*flattening*) der Proteinstruktur (Talmon, 1983).

Bei der Herstellung der Probe ging ich wie folgt vor: Auf ein Stück Parafilm wurde ein Tropfen der Proteinlösung von ca. 5 µl aufgetragen. Zusätzlich wurden

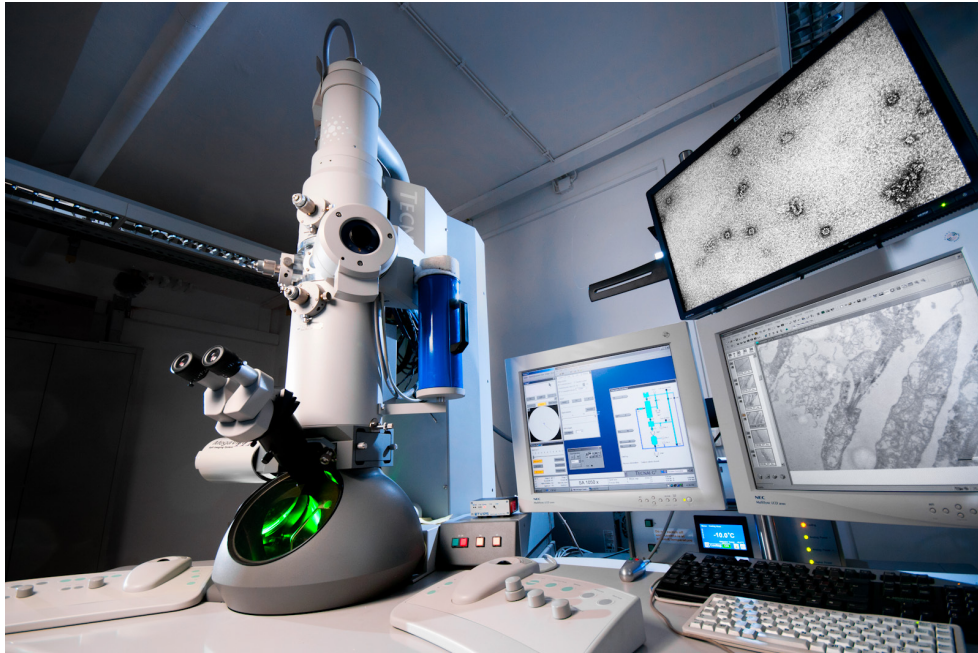


Abb. B.8: Das Tecnai G² 12 BioTWIN Elektronenmikroskop am Institut für Zoologie der JGU Mainz. (Bild © Christoph Kühne)

drei 20 µl-Tropfen destilliertes Wasser und ein 5 µl-Tropfen Kontrastmittel auf dem Parafilm platziert.

Danach wurde die Probe mit dem Kohlefilm des zuvor glimmentladenen Objektträgers aufgenommen. Der Tropfen verteilte sich gleichmäßig und die Proteine bindeten an der Oberfläche. Nun wurde die Probe mit dem destillierten Wasser gewaschen, indem man nacheinander die drei Tropfen mit dem Kohlefilm des Objektträgers berührte.

Daraufhin wurde die überschüssige Flüssigkeit mit einem Stück Filterpapier (z. B. Whatman No. 1) abgenommen und direkt im Anschluss das Kontrastmittel mit dem Kohlefilm aufgenommen. Nach ca. 10 Sekunden Einwirkzeit wurde das Kontrastmittel ebenfalls mit Filterpapier abgesaugt. Der so präparierte Objektträger trocknete dann im Anschluss für ca. zehn Minuten und konnte sogleich im EM verwendet werden.

Als Kontrastmittel kam 2% Uranylacetat oder 2% Uranylformiat zum Einsatz, bezogen von Chemos GmbH (Regenstauf, Deutschland). Eine Anleitung zur Herstellung befindet sich im Anhang auf Seite 133. Die Proben lagerten bei Raumtemperatur und möglichst geringer Luftfeuchtigkeit.

Kryo-Fixierung von Proteinen

Bei der Kryo-Fixierung wurde der Puffer und die enthaltende Probe vitrifiziert (Dubochet et al., 1988). Dieses Vorgehen hat im Vergleich zur Negativkontrastierung verschiedene Vorteile: Die erreichbare Auflösung ist nicht mehr durch ein Kontrastmittel beschränkt, weiterhin befindet sich das Protein in einem möglichst physiologischen Puffer und wird durch den schockartigen Einfrierprozess in seiner natürlichen Konformation fixiert. Ein Nachteil der Methode ist der geringe Phasenkontrast der Proteine im Eis, da sie im Vergleich zum umgebenden Medium nur einen geringen Dichteunterschied aufweisen.

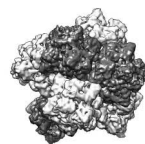
Vor der eigentlichen Probenpräparation müssen die Objektträger aufgrund des Herstellungsprozesses gereinigt werden. Dazu wurden sie in einer Petrischale mit Ether überschichtet und auf einem Laborschüttler für etwa eine Stunde gereinigt. Danach wurde der Prozess mit Aceton wiederholt.

Die eigentliche Präparation geschah mithilfe des Cryoplunger™3 (siehe Seite 20), der den Einfrierprozess weitestgehend automatisierte. Der gewaschene und per Glimmentladung vorbereitete Objektträger wurde mit einer Pinzette in das Gerät eingespannt. Mit einer Pipette wurden 3,5 µl Probe aufgetragen, die sich in der kontrollierten Atmosphäre innerhalb des CP3 akklimatisierten. Nun wurden die überschüssige Probe mit einem Filterpapier (z. B. Whatman No. 1) abgesaugt und der Objektträger in flüssiges Ethan eingetaucht. Das flüssige Ethan produzierte dabei keine isolierende Gasschicht zwischen Ethan und Objektträger, somit konnten sich keine Eiskristalle ausbilden und die Probe wird vitrifiziert (Adrian et al., 1984). Danach wurden die Objektträger in flüssigen Stickstoff gelagert, um ein Auftauen und die Bildung von Eiskristallen, die die Proteine schädigen würden, zu verhindern.

Der Schlüsselschritt des Vorgangs war die Dauer des Absaugens. Hier muss die richtige Dauer in Millisekunden durch Versuche ermittelt werden, um eine möglichst dünne Eisschicht zu erzeugen. Die weitere Handhabung der Proben geschah vollständig unter flüssiger Stickstoffatmosphäre.

3.4 Elektronendetektion

Elektronen sind für das menschliche Auge nicht direkt sichtbar. Um sie nach der Interaktion mit den Proteinen zu visualisieren, haben sich in der Praxis mehrere



Systeme mit Vor- und Nachteilen etabliert. Die älteste Technik mit Negativfilmen ist auch heute noch weit verbreitet, wird aber immer stärker von modernen Digitalkameras abgelöst. Die erst vor Kurzem entwickelte Technik der direkten Elektronendetektion (Bammes et al., 2012) stand mir für meine Versuche in Mainz noch nicht zur Verfügung.

Filme

Ein elektronenempfindlicher Film besitzt feine Silberhalogenkörner (4 bis 5 μm Korngröße) in einer Emulsion über einer Trägerfolie. Eintreffende Elektronen ionisieren diese Halogene zu Silber. In einem anschließenden Entwicklungsprozess werden die Silberpartikel fixiert und der Film kann mithilfe eines Scanners digitalisiert werden.

Der Vorteil von Negativen liegt in ihrer sehr hohen Effizienz detektierbarer Menge (*detective quantum efficiency*) (DQE), in diesem Falle an Elektronen, was für strahlungsempfindliche biologische Proben besonders wichtig ist. Weiterhin kann durch die Abmessungen des Negativs ein großer Bereich aufgenommen werden. Ebenfalls lässt sich eine hohe Auflösung erreichen, die durch die Korngröße der Silberhalogene definiert wird. Für das Vakuumsystem des EM entstehen Nachteile durch die stark hydrophilen Filme und in der Weiterverarbeitung der Aufnahmen, denn zur Digitalisierung wird ein Scanner mit sehr hoher Auflösung benötigt.

Der in dieser Arbeit verwendete Negativfilm war Kodak *electron image film* SO-163 8,3 cm \times 10,2 cm. Das Entwickeln der Negativfilme ist im Anhang auf Seite 133 beschrieben.

Scanner

Die Negativfilme wurden mit einem Primescan D 7100 Trommelscanner der Firma Heidelberger Druckmaschinen (Heidelberg, Deutschland) digitalisiert. Um ein optimales Scanergebnis zu erreichen, wurden die Filme mit einer dünnen Schicht *Mounting Oil* der Firma SDS AG (Norderstedt, Deutschland) auf die Scannerwalze aufgeklebt. Dies soll die Entstehung von Newton-Ringen vermeiden, die das digitalisieren erschweren. Als Scannsoftware diente Linocolor der Firma Heidelberger Druckmaschinen oder Silverfast der LaserSoft Imaging AG (Kiel,

Deutschland). Die Abtastgenauigkeit dieses Trommelscanners liegt bei maximal 11 000 Bildpunkte pro Inch (*dots per inch*) (DPI) und ist frei wählbar.

Digitalkamera

Bei einer für das EM geeigneten Digitalkamera werden die Elektronen mithilfe eines Szintillators in sichtbares Licht umgewandelt und dieses von der Kamera detektiert. Der große Vorteil bei der Verwendung dieser Kameras gegenüber Film ist die sofortige Einsicht in die Ergebnisse. Es wird keine aufwändige Entwicklungsarbeit mehr benötigt und bei eventuell fehlerhaften Proben kann direkt reagiert werden. Für die Automatisierung spielen Digitalkameras eine noch größere Rolle, da die Daten direkt im richtigen Format für die weitere Prozessierung vorliegen. Ein Nachteil der Digitalkameras ist ein geringerer DQE als Film. Dies wird aber in der Regel durch eine größere Zahl an Aufnahmen pro Zeiteinheit wieder ausgeglichen.

Die negativkontrastierten Aufnahmen wurden alle mit einer TemCam-F416 Digitalkamera der Firma TVIPS (München, Deutschland) aufgenommen. Hierbei handelt es sich um eine 4096 x 4096 Pixel *complementary metal oxide semiconductor* (CMOS) Kamera mit einer Pixelgröße von 16 μm . Durch die *binning*-Funktion können vier Pixel zu einem 32 μm messenden Gesamtpixel zusammengefasst werden und somit die Empfindlichkeit (Signal-Rausch-Verhältnis (*signal to noise ratio*) (SNR)) der Kamera erhöhen. Dies lässt sich um eine weitere Stufe wiederholen und insgesamt 16 Pixel (64 μm Kantenlänge) zu einem zusammenfassen. Hierdurch steigt das SNR, allerdings leidet die Auflösung unter dem Prozess.

4 Computer

4.1 Rechen-Cluster

Das Rechen-Cluster, bezogen von der Firma Transtec (Tübingen, Deutschland), bestand aus insgesamt 27 Rechenknoten mit unterschiedlichen Prozessoren. Aus der Zusammenstellung von 36 AMD Octa-Core, 32 AMD Dual-Core und zwei Intel Xeon Quad-Core Prozessoren ergab sich eine Gesamtzahl von 360 Rechenkernen, die jeweils mit 2 Gigabyte (GB) Arbeitsspeicher ausgestattet sind. Die Datensicherung übernahm ein Netzwerkspeicher (*network attached storage*) (NAS)



mit 30 Terabyte (TB) Speicherkapazität. Das verwendete Betriebssystem war Ubuntu Linux.

4.2 Arbeitsplatzrechner

Die Abbildungen, Programmierarbeiten sowie das Verfassen der Arbeit wurden an einem Apple MacBook Pro 13" mit 2,4 GHz Intel Core i5 Prozessor und 16 GB Arbeitsspeicher durchgeführt. Das verwendete Betriebssystem war Mac OS X 10.7.5.

5 Software

Eine vollständige Liste der Softwarepakete wird in Tabelle B.2 und B.3 aufgeführt.

Tab. B.2: Verwendete frei erhältliche Softwarepakete

Name	Version	Arbeitsgruppe	Publikation	Institut	Land
Appion	2.2	NRAMM	(Lander et al., 2009)	TSRI	CA USA
Sparx	2.06	SBIC	(Hohn et al., 2007)	TMC	TX USA
EMAN	1.9	NCMI	(Ludtke et al., 1999)	BCM	TX USA
EMAN2	2.06	NCMI	(Tang et al., 2007)	BCM	TX USA
ctffind3	3.5		(Mindell and Grigorieff, 2003)	HHMI	VA USA
Chimera	1.8	RBVI	(Pettersen et al., 2004)	UCSF	CA USA
DireX	0.61	CSB Group	(Wang and Schröder, 2012)	ICS-6	Jülich DE
Gromacs	4.6.3	verschiedene	(Van Der Spoel et al., 2005)	verschiedene	USA, DE, SE
VMD	1.9.1	TCB Group	(Humphrey et al., 1996)	Beckman	IL USA

Tab. B.3: Verwendete kommerziell erhältliche Softwarepakete

Name	Version	Publikation	Firma	Land
IMAGIC	110325	(van Heel et al., 1996)	Image Science	Berlin DE
EM-Menu	4.0.9.23	(Daberkow et al., 1996)	TVIPS	München DE

Appion

Appion ist ein modular aufgebautes, internetbasiertes Programm zum Automatisieren von 3D-Rekonstruktionen (Lander et al., 2009). Das Programm bietet

eine Internetseite mit Formularen an, die spezifische externe Programme ansteuert. Dabei werden die Ergebnisse in einer Datenbank gespeichert, auf die mit verschiedenen Auswertungsskripten zugegriffen werden kann.

Sparx

Für die 3D-Rekonstruktionen in dieser Arbeit wurde bevorzugt das Programm **Sparx** (Hohn et al., 2007) benutzt. Es zeichnet sich aus durch akkurate Algorithmen wie ungleichförmige Fouriertransformation (Penczek et al., 2004) und Elemente aus `cctbx`¹ (Grosse-Kunstleve et al., 2002).

Diese Komponenten werden in einem dreistufigen Programmkonzept zusammengefasst. Die erste Stufe ermöglicht den Zugriff auf jede einzelne Funktion der Programme. Dies passiert in einer `Python`-Programmierumgebung (siehe Abschnitt B.7.1). Die zweite Stufe stellt fertige Algorithmen aus der **Sparx**-Bibliothek zur Verfügung, die in der Kommandozeile ausgeführt werden und an dem Präfix `sx` zu erkennen sind. Als dritte Stufe bietet sich eine grafische Benutzeroberfläche an, die mehrere Funktionen der Software in einem Diagramm verknüpft. Programmiererfahrung ist somit nicht mehr zwingend erforderlich.

Sparx wurde, ausgenommen von zeitkritischen Algorithmen, in der Programmiersprache `Python` entwickelt. Spezielle Konzepte wie „*Userfunctions*“ (siehe Abschnitt B.6.7) ermöglichen eine einfache Erweiterung der Standardalgorithmen.

EMAN

EMAN (Ludtke et al., 1999) ist das Vorgängerprogramm von **EMAN2** und bietet hauptsächlich Kommandozeilenwerkzeuge an. Es bietet eine komplette Programmumgebung für die Einzelpartikelanalyse.

EMAN2

EMAN2 (Tang et al., 2007) ist die Weiterentwicklung von **EMAN** und stellt neben den Kommandozeilenbefehlen auch eine GUI zur Verfügung. Durch die gemeinsame Verteilung mit dem Programmpaket **Sparx** werden hier viele Module gemeinsam

¹Einer Sammlung von Algorithmen für Kristallographie



benutzt. Die EMAN2-Module kann man an dem Präfix `e2` in der Kommandozeile erkennen.

IMAGIC

IMAGIC (van Heel and Keegstra, 1981; van Heel et al., 1996) ist ein Softwarepaket, das im Vergleich zu EMAN, EMAN2 oder Sparx weniger auf vollständige Automatisierung ausgelegt ist und in dem viele Parameter frei gewählt werden können. Bei Bedarf können komplexere Abläufe in einer Skriptdatei zusammengefasst und ausgeführt werden.

ctffind3

ctffind3 (Mindell and Grigorieff, 2003) ist ein Programm zum Berechnen des Defokus und des Astigmatismus einer elektronenmikroskopischen Aufnahme. Die gemessenen Werte werden für die CTF-Korrektur benötigt.

EM-Menu

Die Software EM-Menu dient zum Steuern elektronenmikroskopischer Digitalkameras. Es wurde am Tecnai G² 12 BioTWIN (siehe Kapitel B.3.2) TEM benutzt.

Chimera

Bei UCSF Chimera (Goddard et al., 2007; Pettersen et al., 2004) handelt es sich um eine 3D-Visualisierungssoftware, spezialisiert auf den Einsatz im Bereich Strukturbiologie. Chimera bietet eine Programmierschnittstelle in Python und ist somit durch eigene Programme erweiterbar.

Direx

Direx (Schröder et al., 2007) stellt mehrere kommandozeilenbasierte Programme zum Einpassen atomarer Strukturen in eine Elektronendichtekarte zur Verfügung.

Gromacs

Gromacs (Van Der Spoel et al., 2005; Hess et al., 2008) ist ein Programm zur Moleküldynamik-Simulation.

VMD

VMD (Humphrey et al., 1996) ist wie **Chimera** eine 3D-Visualisierungssoftware mit einem stärkeren Fokus auf Moleküldynamik-Simulationen. Diese Software stellt einen speziellen Grafikkern zur Verfügung, der Atom-Trajektorien schneller verarbeiten kann. Bei diesen Atom-Trajektorien wird die Position für jedes einzelne Atom des Moleküls im 3D-Raum in einer zeitlichen Abfolge gespeichert. Somit ergeben sich für die jeweiligen Atome Pfade im Raum.

Latex und Tex

T_EX ist ein Textsatzsystem zum Erstellen von Textdokumenten. Für diese Doktorarbeit wurde **L^AT_EX** benutzt, welches das rudimentäre **T_EX** mit weiteren Makros zur einfacheren Verwendung erweitert. Durch dieses Textsatzsystem muss der Benutzer sich z. B. nicht um die Nummerierung von Grafiken kümmern und das ganze Dokument besitzt ein einheitliches Layout.

DokuWiki

Zu Dokumentationszwecken der Arbeitsgruppe wurde ein DokuWiki (Gohr, 2013) erstellt. In diesem Wiki wurden selbstverfasste Handbücher veröffentlicht, die jedes Gruppenmitglied erweitern und verbessern kann. Durch das Bearbeiten von verschiedenen Mitarbeitern wurden die Anleitungen unmissverständlicher und zu einer häufig genutzten Quelle für neue Gruppenmitglieder.

6 Bildverarbeitende Methoden

Die Grundidee der Einzelpartikelanalyse besteht darin, dass man ein 3D-Modell im Raum anhand von nur drei Projektionen des gleichen Objektes vollständig beschreiben kann (Crowther et al., 1970b,a; van Heel, 1987). Aufgrund eines



niedrigen Signal-Rauschen-Verhältnisses (*signal to noise ratio*) (SNR) und durch Bildfehler werden in der Praxis eine viel höhere Anzahl an Projektionen benötigt.

In diesem Abschnitt wird auf die genaue Vorgehensweise eingegangen, wie aus EM-Aufnahmen (Abb. B.9) eine 3D-Struktur berechnet wird. In Abb. B.10 wird dieser Arbeitsablauf in einem Diagramm dargestellt. Die Liste der verwendeten Module befindet sich im Abschnitt B.6.14.

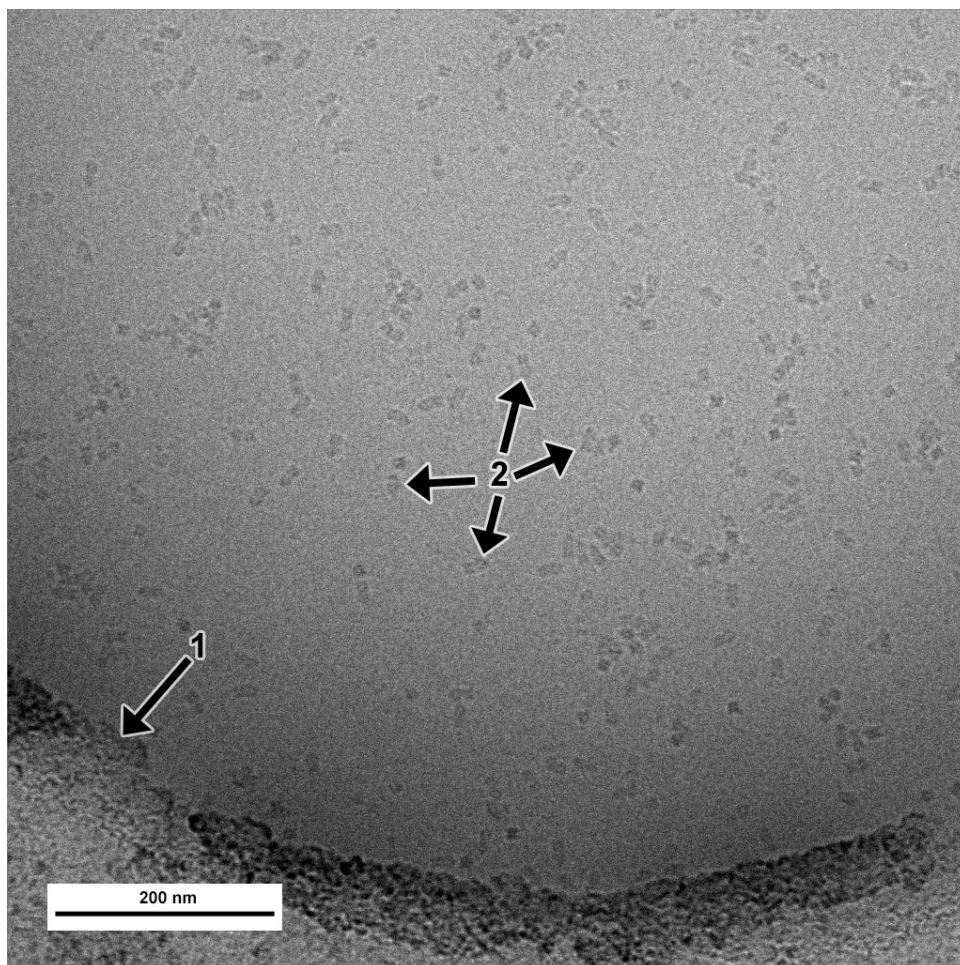


Abb. B.9: Elektronenmikroskopische Aufnahme des *Astacus leptodactylus*-Hämocyanins in amorphem Eis. (1) zeigt deutlich den Rand des Kohlefilms; (2) zeigt die einzelnen Hämocyanin Partikel.

Erstellt wurde die Aufnahme an einem Tecnai G² 12 BioTWIN und der TemCam-F416 Digitalkamera. Die nominelle Vergrößerung betrug 71 540 \times bei einem Defokus von 3,3 μ m.

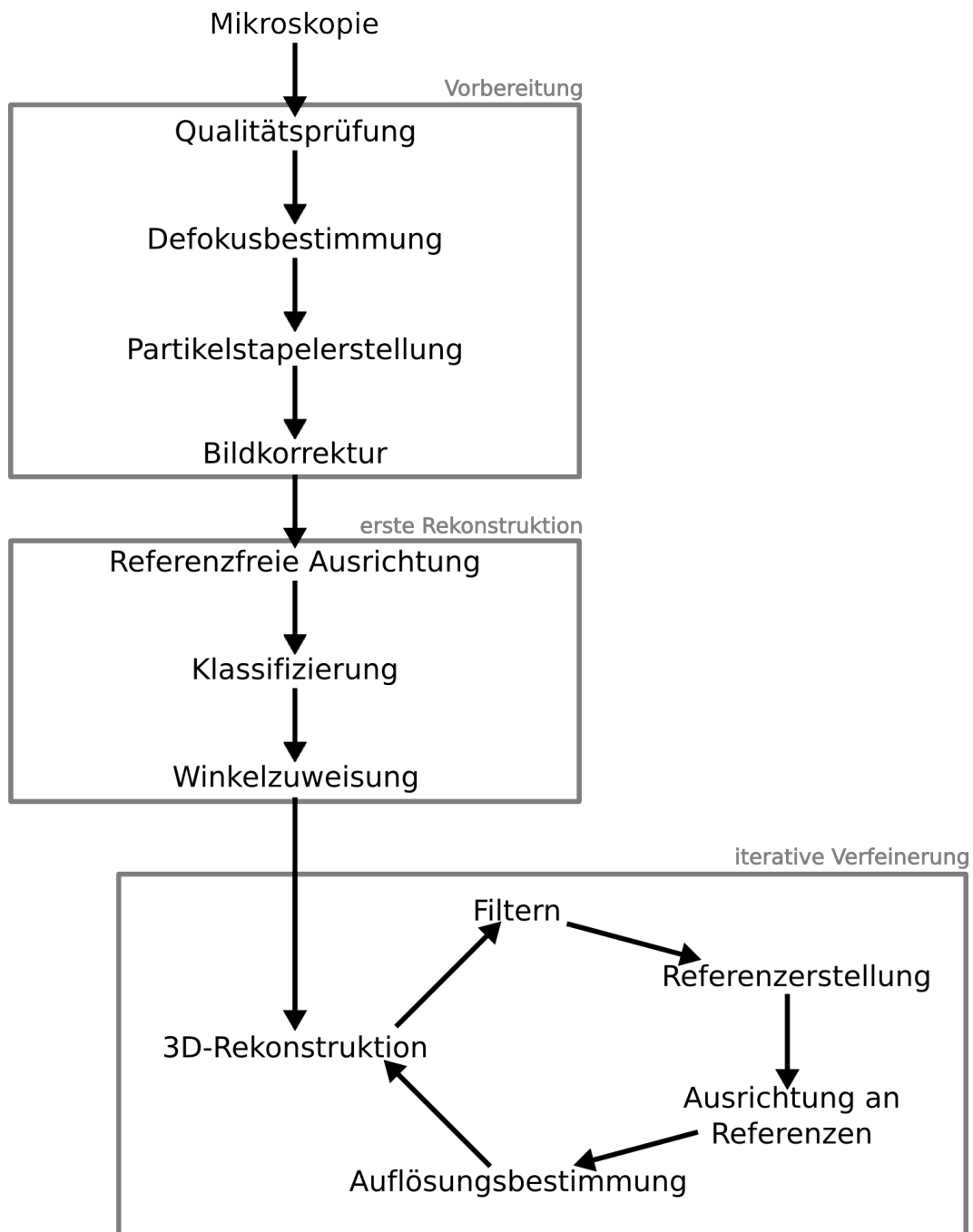


Abb. B.10: Arbeitsablauf von der Elektronenmikroskopie zur fertigen 3D-Rekonstruktion.



6.1 Automatische Stapelverarbeitung mit Appion

Appion wurde in dieser Arbeit hauptsächlich zur Erstellung der Partikelstapel benutzt (siehe Abbildung B.10 „Vorbereitung“). Da bei der 3D-EM hunderte von Aufnahmen angefertigt und nach einem ähnlichen Algorithmus analysiert wurden, bietet sich eine Automatisierung der verschiedenen Schritte an (Potter et al., 1999; Carragher et al., 2000). Appion speichert alle benötigten mikroskop-, aufnahme- und partikelspezifischen Parameter in einer Datenbank, auf die mit Appion verknüpfte Programme zurückgreifen können.

Qualitätskontrolle

Um die Qualitätsprüfung zu beschleunigen bietet Appion eine spezielle Bildergalerie an, in der man die elektronenmikroskopischen Aufnahmen analysieren und bei Bedarf vor der weiteren Prozessierung ausschließen kann. Dabei werden die Bilder auf die Eisqualität (Abb. B.11) sowie Partikelverteilung (Abb. B.12) überprüft. Wenn die Moleküle sich im Eis berühren, sind sie zur weiteren Verarbeitung nicht geeignet.

Parallel wird eine Schnelle Fourier-Transformation (*fast Fourier transform*) (FFT) berechnet, da Bildfehler wie Bewegungsunschärfe hiermit besser zu erkennen sind (Abb. B.13). Die Aufnahmen verlieren dabei Detailinformation in Richtung der Bewegung und die Moleküle verformen sich visuell. In der FFT sind diese Bewegungen durch unvollständige Thon-Ringe (Thon, 1966) zu erkennen (Abb. B.13 B).

Um die Bewegungsunschärfe einzuschränken wird mit neuen Objektträgermaterialien wie Siliciumcarbid geforscht (Yoshioka et al., 2010), um die Materialbewegung im Objektträger zu reduzieren. Neuere Untersuchungen mithilfe von direkter Elektronendetektion versuchten die Bewegung in einer späteren Bildkorrektur herauszurechnen (Li et al., 2013; Veesler et al., 2013). Hier wurden Filme mit etwa 10 Bildern pro Sekunde aufgenommen und die Einzelbilder aneinander ausgerichtet. Somit kann die Bewegungsunschärfe durch Bewegung des Objektträgers fast vollständig eliminiert werden. Leider stand mir diese Technik nicht zur Verfügung.

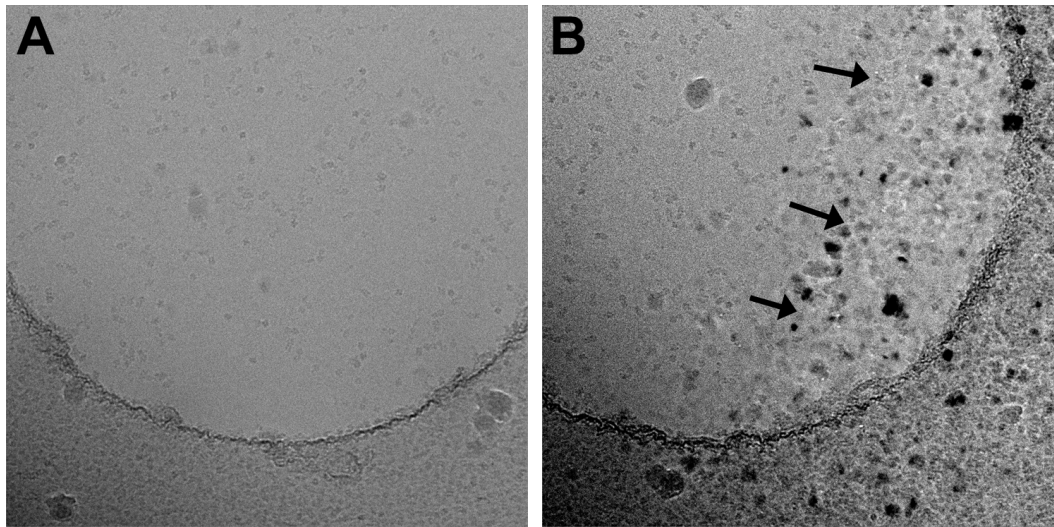


Abb. B.11: Aufnahmen mit kristallinen Bereichen müssen aussortiert werden, da die Eiskristalle Moleküle beschädigen können. (A) Aufnahme mit amorphem Eis. (B) Aufnahme mit Eiskristallen (Pfeile).

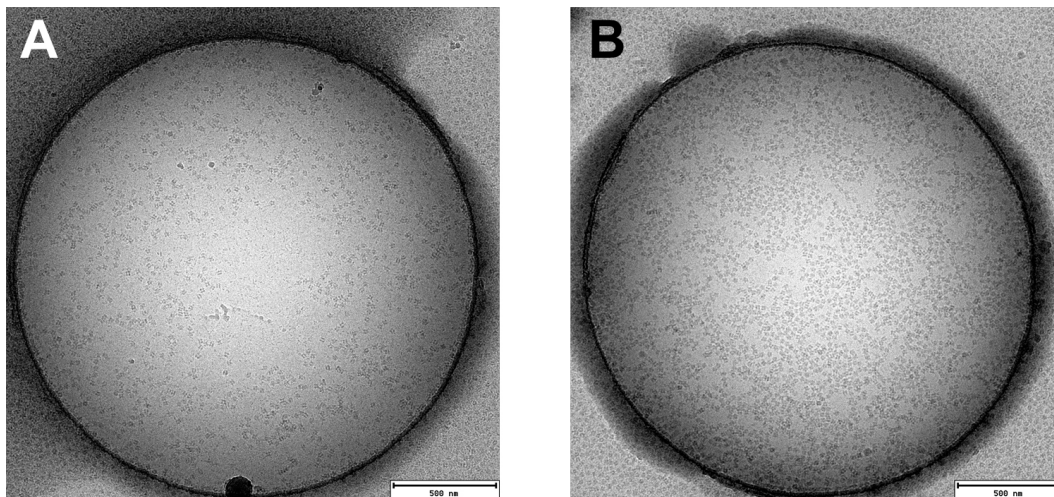


Abb. B.12: Verteilung der Moleküle im vitrifizierten Eis: (A) Bevorzugte Verteilung der einzelnen Moleküle im Eis, die Proteine sind in hoher Konzentration vorhanden ohne sich zu berühren. (B) Hohe Proteinkonzentration, die Moleküle berühren sich untereinander und sind somit auszuschließen.

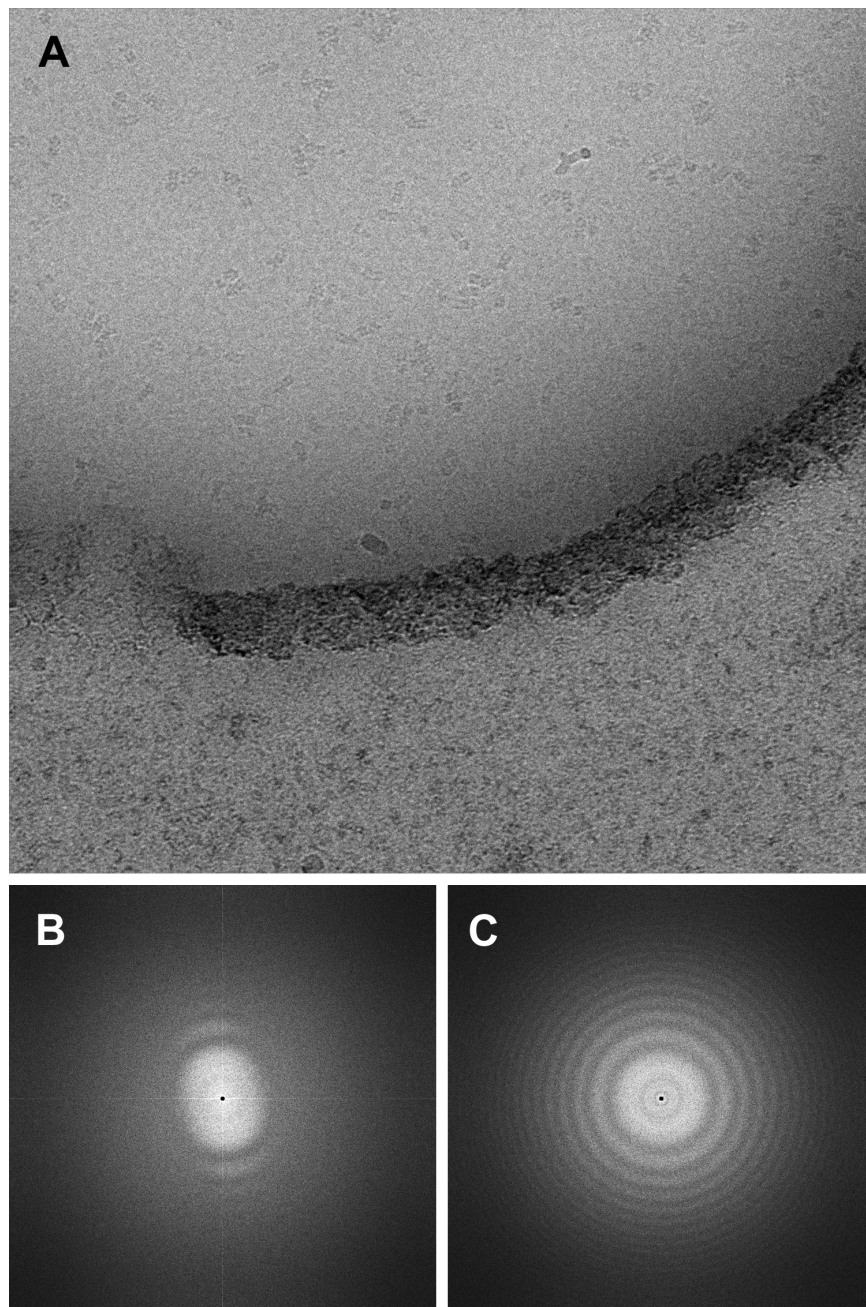


Abb. B.13: Analyse der Bewegungsunschärfe bei EM-Bildern: (A) In einer EM-Aufnahme lässt sich die Bewegungsunschärfe oft nicht direkt erkennen. (B) In der FFT wird die fehlende Information aufgrund der unvollständigen Thon-Ringe in Bewegungsrichtung deutlich. (C) Eine FFT ohne Bewegungsunschärfe, die Ringe sind vollständig.

Bestimmung des Defokus

Wie im Abschnitt B.3.1 beschrieben, haben biologische Proben im TEM nur sehr wenig Kontrast und sind im fokussierten Zustand nicht zu erkennen (Zernike, 1953). Die Aufnahmen wurden aus diesem Grund im Defokus erstellt. Um die Bildfehler durch das Defokussieren herauszurechnen, muss der Wert des Defokus der Aufnahme möglichst genau bestimmt werden. *Appion* benutzt hier die Programme *ctffind3* (Mindell and Grigorieff, 2003) oder *ACE* (Mallick et al., 2005).

Da sich der Wert des Defokus im Durchmesser der Thon-Ringe der FFT widerspiegelt, lässt sich dieser Wert in Verbindung mit den Mikroskopparametern aus der *Appion*-Datenbank berechnen. Die Programme erstellen Diagnosebilder (Abb. B.14), die der Benutzer auf Fehler untersuchen muss. *ctffind3* zeigt in den Diagnosebildern auf der rechten Hälfte die FFT des EM-Bildes und versucht auf der linken Hälfte die FFT mit dem berechneten Defokuswert zu simulieren. Sind die beiden Seiten ungleich, müssen die Bilder vor der weiteren Prozessierung ausgeschlossen werden.

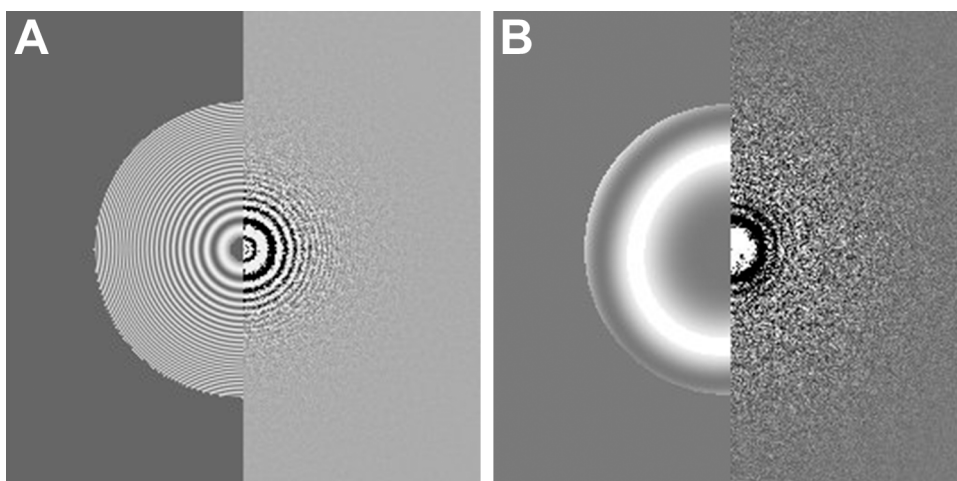


Abb. B.14: Diagnosebilder aus *ctffind3* zeigen auf der rechten Seite die FFT des EM-Bildes und auf der linken Seite die simulierte FFT. (A) Zeigt eine gelungene Simulation. (B) Die Defokusparameter wurden fehlerhaft berechnet.



Partikelstapelerstellung

Für die Einzelpartikelanalyse werden quadratische Bildausschnitte benötigt, die idealerweise nur einen Partikel pro Bild enthalten. Dazu werden die Partikel innerhalb der elektronenmikroskopischen Aufnahme selektiert und die genauen Koordinaten gespeichert (Abb. B.15).

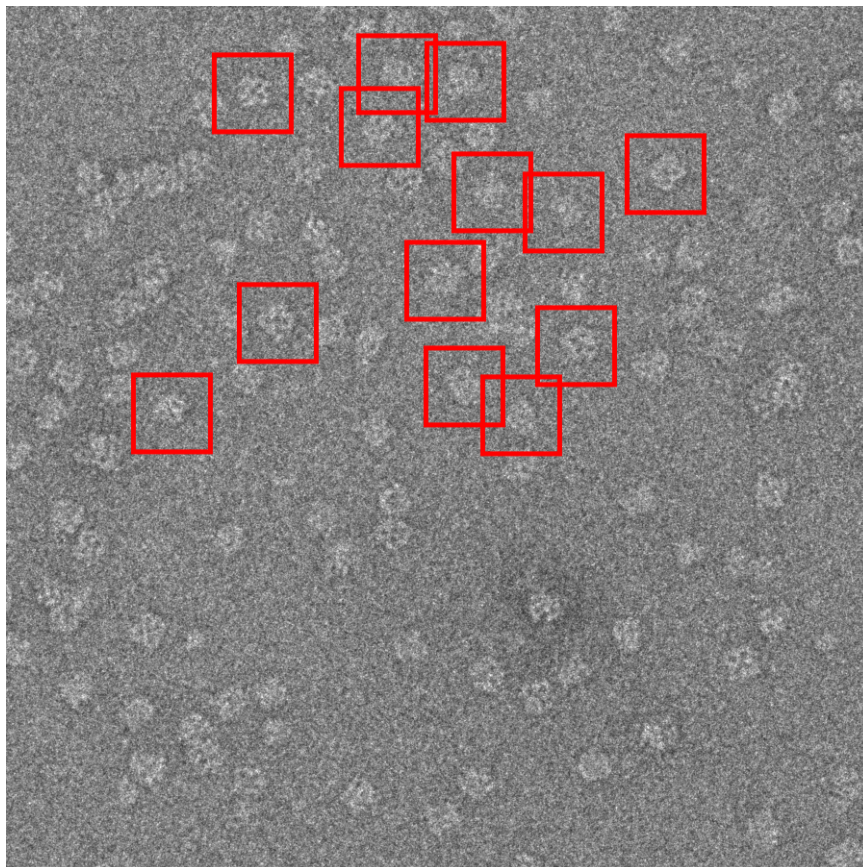


Abb. B.15: Manuelle Selektion des negativ kontrastierten *D. pulex* Hämoglobin. Die selektierten Partikel sollen sich nach Möglichkeit nicht berühren.

Die Selektion der Partikel kann wahlweise von Hand oder mit verschiedenen Algorithmen (Roseman, 2004; Mallick, 2004; Glaeser, 2004; Chen and Grigorieff, 2007; Voss et al., 2009) automatisch durchgeführt werden. Dabei ist die manuelle Auswahl oft sehr zeitintensiv, aber meist qualitativ hochwertiger, da jeder Partikel visuell überprüft wird. Die automatischen Methoden können viele Partikel in kurzer Zeit selektieren, müssen aber von *false positives*² befreit werden.

²Falsche Partikel, die der Algorithmus als richtig einstuft

Nach der Selektion der Partikel wurden diese aus der Aufnahme ausgeschnitten und in einer Datei als Partikelstapel gespeichert. Die quadratische Box sollte etwa dem 1,5- bis 2-fachen der maximalen Molekülabmessung entsprechen (Downing and Glaeser, 2008; Ludtke, 2012), um die Bildkorrektur nicht negativ zu beeinflussen.

6.2 Bildkorrektur des Partikelstapels

Wie im Abschnitt B.3.1 beschrieben, wurde bei biologischen Proben das Bild defokussiert, um einen Kontrast zu erhalten. Hier ergab sich durch die Kontrast-Transfer-Funktion (*contrast transfer function*) (CTF) eine Änderung der Intensität von hell bis dunkel, abhängig von den Raumfrequenzen der Aufnahme (Abb. B.16 und Abb. B.5 auf Seite 19). Um die Intensitäten zu korrigieren, muss die CTF möglichst genau berechnet werden, indem die nötigen Parameter (sphärische Aberration C_s , Defokus Δf , Beschleunigungsspannung U_B etc.) aus der **Appion**-Datenbank geladen werden. Für die eigentliche Bildkorrektur wurde der jeweilige Partikel im Fourierraum mit der CTF multipliziert (Zhu et al., 1997). Dadurch wurden die negativen Halbwellen positiv und der Partikel interpretierbar (Abb. B.16 B). Um die verbleibenden Nullstellen (kein Kontrast) der CTF auszugleichen, müssen die elektronenmikroskopischen Aufnahmen in einem Defokusbereich zwischen etwa $0,7 \mu\text{m}$ bis $4,0 \mu\text{m}$ angefertigt werden. Durch die anschließende Mittelung besaßen fast alle Frequenzbereiche Kontrast (Abb. B.16 C).

Bei Verwendung der Software **Sparx** wurde ein korrigierter und ein unkorrigierter Partikelstapel erzeugt. Dem unkorrigierten wurden die nötigen Parameter für die CTF-Korrektur in den *Header* der Datei eingetragen. Innerhalb von **Sparx** besteht die Möglichkeit die Korrektur für eine Gruppe von Partikeln gleichzeitig zu berechnen; dieses Vorgehen reduziert Rechenfehler bei stark verrauschten Einzelbildern (Zhu et al., 1997). Diese Funktion wurde in **Sparx** in den finalen Iterationen aktiviert.

6.3 Referenzfreie Ausrichtung

Um sodann die einzelnen Partikel analysieren zu können, mussten diese im Bildausschnitt zentriert werden. Zur Ausrichtung wurde immer eine Referenz benötigt, zu der die einzelnen Partikel ausgerichtet wurden. Aufgrund des starken

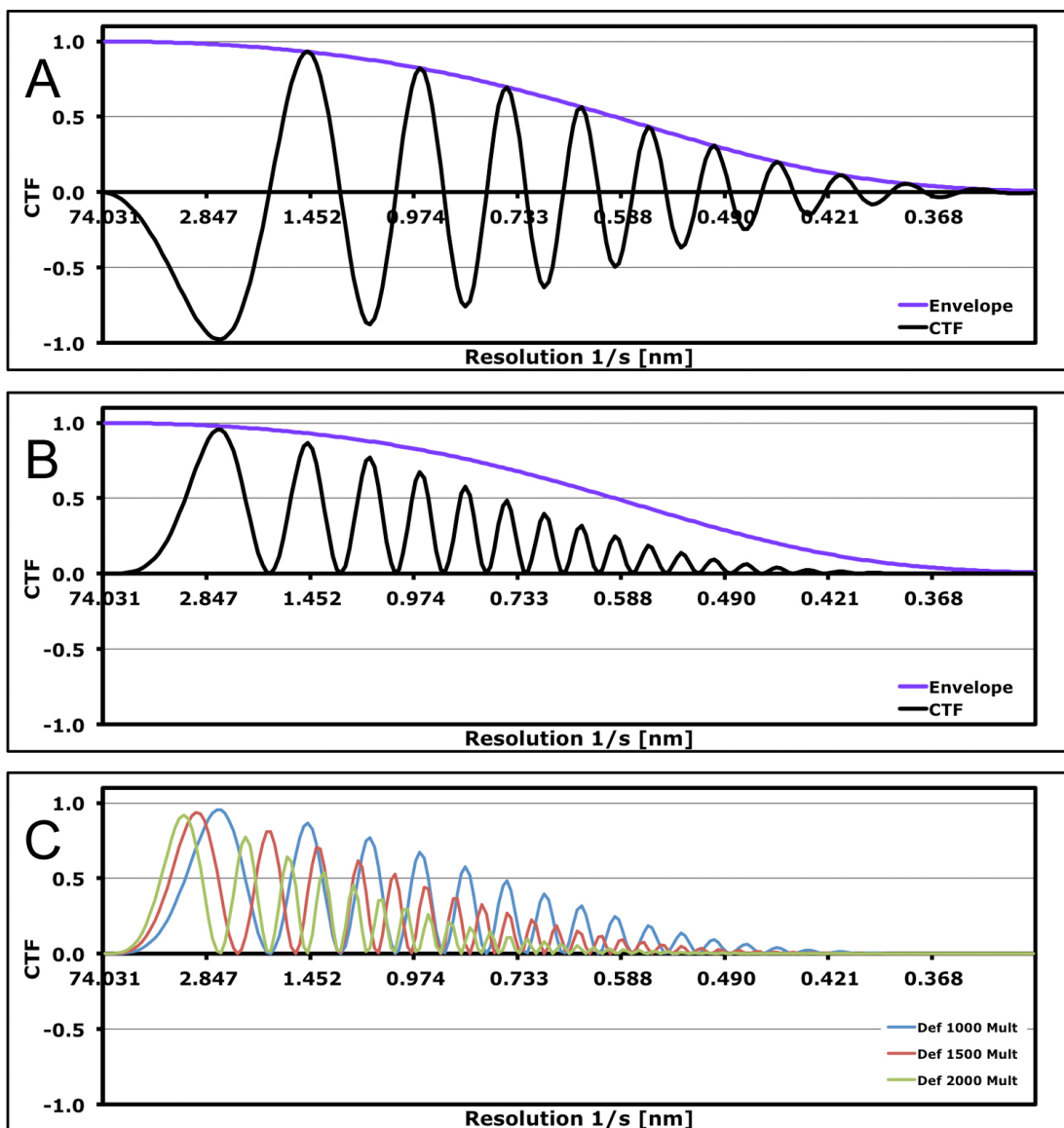
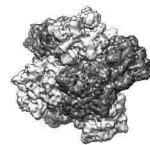


Abb. B.16: A) Beispielhaft eine CTF von einem Tecnai G² 12 BioTWIN bei einem Defokus von 1,0 μm . B) Nach der Korrektur (CTF^2). C) Mehrere Aufnahmen mit unterschiedlichen Defokus füllen die Nullstellen auf (Defokus von 1,0 μm , 1,5 μm und 2,0 μm).

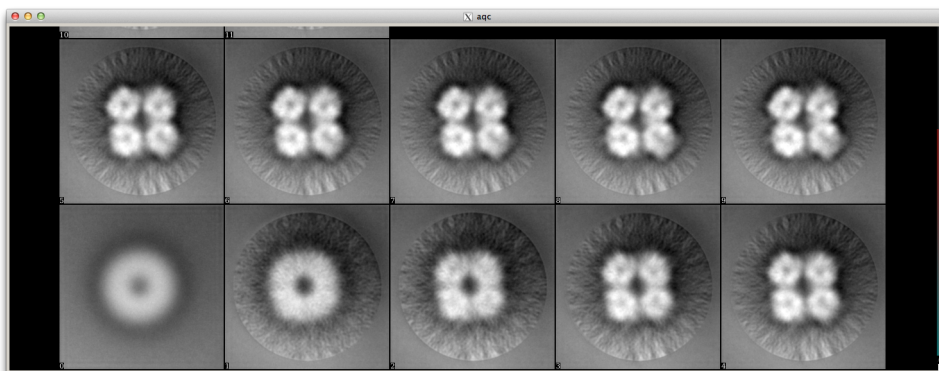


Abb. B.17: Entwicklung der Referenzen des *Limulus polyphemus* Hämocyanin Datensatzes bei referenzfreier Ausrichtung. Das erste Bild links unten entspricht dem Mittel des gesamten Datensatzes durch Aufaddieren der Einzelpartikel. Nach rechts und oben hin sind die jeweiligen Referenzen nach jeder Iteration zu sehen. Die Partikel wurden schrittweise an mehr Details ausgerichtet.

Rauschens (siehe Abschnitt B.6.4) in EM Aufnahmen, können sich hier bei einer falschen Referenz Fehler einschleichen und das Endergebnis negativ beeinflussen (Penczek et al., 1992). Aus diesem Grund wurde eine Methode benutzt, die ohne Einfluss des Benutzers eine Referenz erzeugte (*reference free alignment*). Hierzu wurden alle Einzelbilder im Stapel zu einem einzigen Bild aufaddiert. Dies stellte zu Anfang einen Mittelwert aus allen Partikeln im Datensatz dar und erschien in der Regel als ein zentrierter diffuser weißer Kreis. Diese Abbildung wurde als Referenz für die weitere iterative Ausrichtung (siehe Kapitel B.6.9) verwendet. Abbildung B.17 zeigt die so erstellten Referenzen für das Hämocyanin von *Limulus polyphemus*.

6.4 Klassifizierung

Elektronenmikroskopische Aufnahmen von biologischen Proben sind stark ver- rauscht, da die in Eis eingebetteten Moleküle eine ähnliche Dichte wie das umge- bende Medium haben und somit der Amplitudenkontrast nur ca. 10% beträgt (Penczek, 2010b). Weiterhin sind vitrifizierte biologische Proben empfindlich in Bezug auf die Elektronendosis. Typische Maximalwerte dürfen 15 bis 20 Elektro- nen pro Quadrat-Ångstrom nicht überschreiten, da die Proben sonst beschädigt werden (Adrian et al., 1984; Dubochet et al., 1988). Bei den so entstandenen Bil-



den muss das Signal-Rausch-Verhältnis (*signal to noise ratio*) (SNR) verbessert werden, um den Informationsgehalt zu erhöhen. Um dies zu erreichen, wurden die Einzelpartikel mit gleicher Orientierung in einer Klasse zusammengefasst und addiert (Klassensummen). Die eigentliche Bildinformation addierte sich schneller als das Rauschen (Schatz, 1992). Abbildung B.18 zeigt den Effekt mit 1 200 zufällig verrauschten Bildern von *Daphnia pulex*.

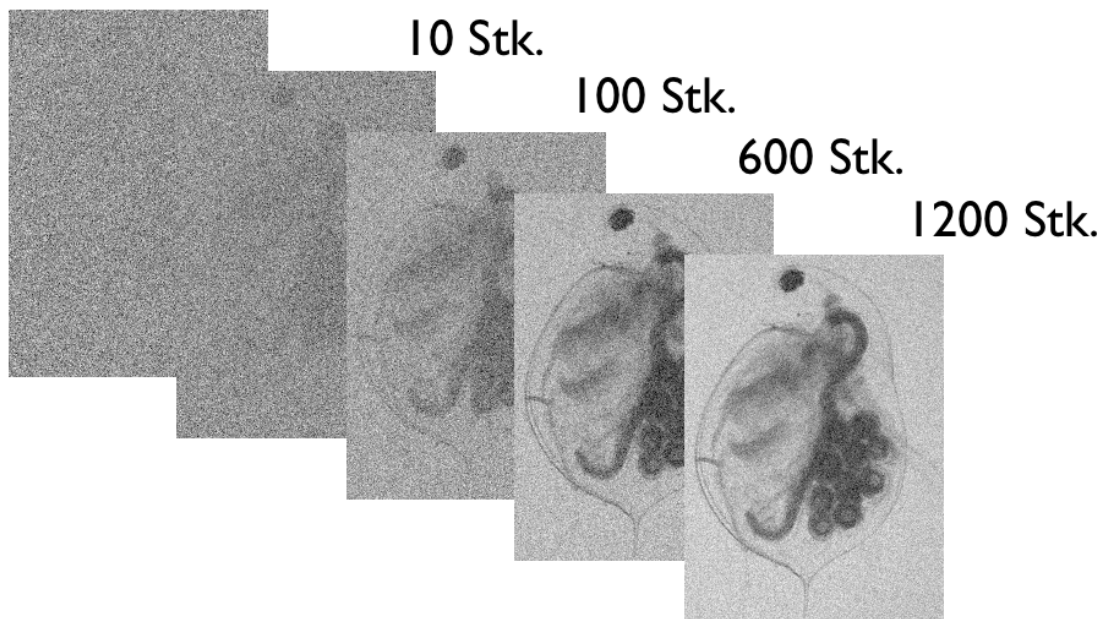


Abb. B.18: Hier wird ein stark verrauschtes Bild von *Daphnia pulex* (links oben) nacheinander mit weiteren zufällig verrauschten Bildern des gleichen Objektes aufaddiert. Bei 1 200 Bildern ist das Rauschen deutlich reduziert. Mit diesem Effekt werden Klassensummen erzeugt. (© Originalbild: wikipedia, Paul Hebert)

Dank der referenzfreien Ausrichtung wurden die einzelnen Partikel zentriert und anhand der dominanten Details ausgerichtet. Die Klassifizierung übernimmt das Modul `sxisax.py` mit der Methode zur iterativen, stabilen Ausrichtung und Klassifizierung (*iterative stable alignment and clustering*) (ISAC). Es handelt sich um eine Erweiterung des K-means Algorithmus (Hartigan, 1975), um stabile und gleich große Klassen (*Equal-size group K-means (EQK-means)*) von Partikeln zu erzeugen (Yang et al., 2012a). Die Partikel wurden dabei per Kreuzkorrelation mit anderen Partikeln verglichen und bei hoher Übereinstimmung gruppiert. Dieser Prozess wird iterativ für eine definierte Anzahl von Runden wieder-

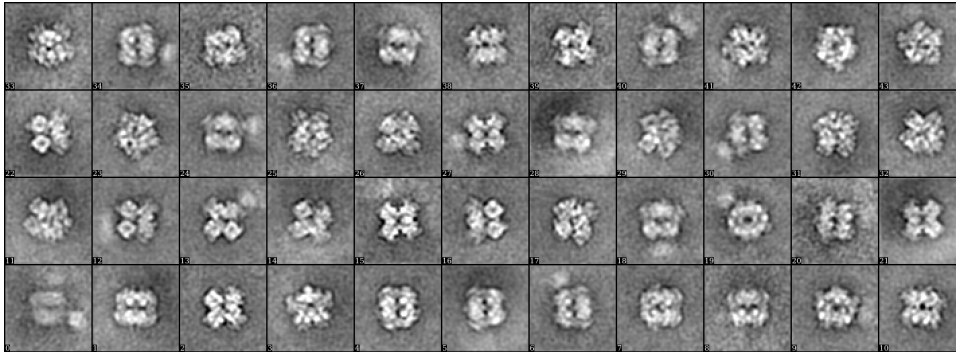


Abb. B.19: Mit der ISAC-Methode erstellte Klassensummen aus einem 16 000 Partikel großen Kryo-EM-Datensatz von *L. polyphemus*-Hämocyanin. Die Klassensummen enthalten zwischen 25 und 40 Einzelpartikel.

holt. Partikel mit nicht reproduzierbaren Klassifikationen wurden ausgeschlossen. Durch den iterativen Prozess wurde vermieden, dass Partikelbilder versehentlich in eine Gruppe gelangen bzw. dort verbleiben. Das Ergebnis (Abb. B.19) waren vergleichbar große Gruppen von Partikelbildern die einen wählbaren Pixelfehler nicht übersteigen (Yang et al., 2012a). Das Verfahren kann mehrfach für die ausgeschlossenen Partikel wiederholt werden bis die gewünschte Zahl an Klassensummen vorhanden ist.

6.5 Winkelzuweisung

Um aus 2D-Projektionen ein 3D-Objekt zu errechnen, muss die Orientierung der 2D-Projektionen im Raum ermittelt werden, bzw. die Orientierung relativ zueinander, dies geschah mit der *common line*-Methode. Es basiert auf dem *central section theorem*, welches besagt, dass alle 2D-Projektionen eines 3D-Objekts im Realraum immer eine 1D-Projektion durch das Zentrum gemeinsam haben (Brewer, 1956; Crowther et al., 1970a). Dabei werden für asymmetrische Objekte mindestens drei Projektionen benötigt, um es vollständig zu beschreiben (van Heel, 1987). Die Orientierung wird in den drei Euler-Raumwinkel α , β und γ sowie den translationalen Verschiebungsvektoren x , y und z angegeben. Anhand dieser lässt sich die Projektionsrichtung genau definieren (Abb. B.20).

Um die Orientierung zwischen den unterschiedlichen ISAC-Klassensummen (1D-Projektionen) zu bestimmen, wurde das Modul `sxfind_struct.py` (Penczek et al., 1996) benutzt. Im Gegensatz zur *angular reconstitution*-Methode (van Heel, 1987),

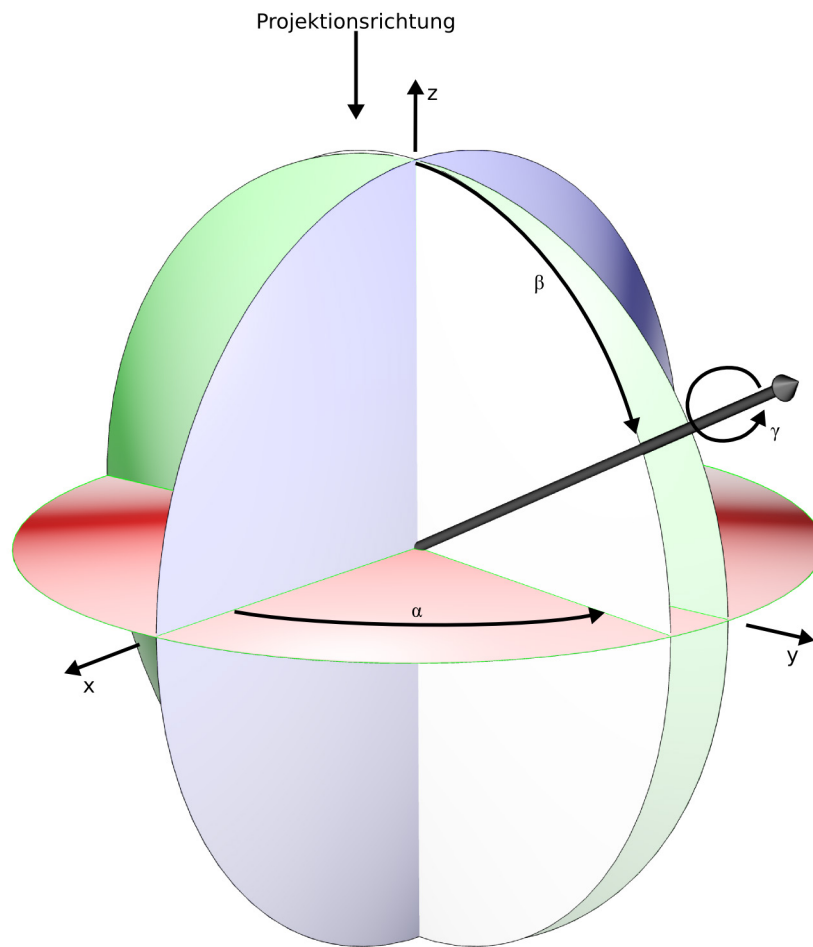


Abb. B.20: Definition der Euler-Raumwinkel α , β und γ sowie translationalen Vektoren x , y und z . Mit der Angabe dieser Werte lässt sich die Orientierung einer Projektion genau bestimmen.

in der die Orientierung für genau drei Projektionen bestimmt wird und weitere Projektionen relativ zu den ersten drei ausgerichtet werden, arbeitet das Modul `sxfind_struct.py` mit möglichst vielen (getestet mit 80) Projektionen. Diese Projektionen wurden mit einer *simultaneous minimization*-Technik gleichzeitig ausgerichtet und iterativ verfeinert, um Fehler zu minimieren (Penczek et al., 1996).

6.6 3D-Rekonstruktion

Moderne Algorithmen zur 3D-Rekonstruktion beruhen auf dem *central section theorem* (DeRosier and Klug, 1968; Crowther et al., 1970a). Dies besagt, dass die 2D-FFT einer Projektion einer Fläche in der 3D-FFT des 3D-Modells entspricht. Durch die Orientierung der Fläche im Raum lässt sich auf die jeweiligen Raumwinkel schließen. Durch Addition möglichst vieler unterschiedlich orientierter Projektionen in diesem Fourierraum erhält man durch inverse FFT die Elektronendichtekarte im Realbildraum.

Da durch die Steigerung der Projektionen die erforderliche Rechenleistung überproportional steigt, wurde weiter nach optimalen Algorithmen zur Rekonstruktion gesucht. Bis heute hat sich der *gridding-based direkt fourier reconstruction*-Algorithmus (GDFR) als einer der genauesten und schnellsten herausgestellt (Zhang et al., 2008) und ist in den Rekonstruktionsprozess von **Sparx** implementiert (Penczek et al., 2004).

Sparx bietet zusätzlich die Möglichkeit, die CTF-Korrektur während der 3D-Rekonstruktion durchzuführen. Hier wurden die unkorrigierten Partikel benutzt und Gruppen mit gleichen CTF-Werten erzeugt, die dann getrennt rekonstruiert wurden. Diese Dichtekarten besaßen, wie die Klassensummen, ein besseres Signal-Rauschen-Verhältnis und das Ergebnis der CTF-Korrektur wurde genauer, da weniger Rauschen verstärkt wird (Penczek et al., 1996). Die einzelnen 3D-Rekonstruktionen wurden danach zu einem Gesamtvolumen addiert.

6.7 Filtern der 3D-Rekonstruktion

Maskieren

Die 3D-Rekonstruktion ist aufgrund der 1,5-2-fach größeren Box deutlich größer als das Protein selbst. Um die Qualität der nachfolgenden 2D-Ausrichtungen zu erhöhen, wurde das Volumen außerhalb des Partikels maskiert. Ein 3D-Volumen (Maske) wurde erzeugt, in dem alle zu maskierenden Bereiche eine Null besaßen, die Bereiche die erhalten werden sollen, eine Eins. Diese Maske wurde dann mit der 3D-Rekonstruktion multipliziert.

Bei der Erstellung der Maske war darauf zu achten, dass keine scharfen Kanten zwischen Null und Eins entstehen, da es sonst zu Artefakten im Fourierraum



```
1 simmap = get_im("simulated_map.hdf") # öffne Datei mit simulierter Dichtekarte
rt = rops_table(simmap) # bestimme Intensitäten der Raumfrequenzen
3 write_text_file([rt], "frequenzen_from_pdb.txt") # schreibe Intensitäten in Datei

5 map2sharp = get_im("3dvolume.hdf") # öffne zu schärfende Datei
fftip(map2sharp) # Fouriertransformieren
7 ro = rops_table(map2sharp) # bestimmen der Intensitäten
rt = read_text_file("frequenzen_from_pdb.txt") # Intensitäten einlesen
9 for j in xrange(len(ro)):
    ro[j] = sqrt(rt[j]/ro[j]) # anpassen der Intensitäten
11 map2sharp = fft(filt_table(map2sharp, ro)) # neu Intensitäten anwenden
drop_image(map2sharp, "map_sharp.hdf") # geschärfte Datei speichern
```

Abb. B.21: *Sharpening* der 3D-Rekonstruktion mit Sparx.

kommen konnte. Das wurde durch die Verwendung des Moduls `proc3d` aus dem EMAN-Softwarepaket vermieden.

Schärfen (*Sharpening*)

Auflösungen, in denen Sekundärstrukturelemente sichtbar werden, besitzen, aufgrund des geringen SNR, hohe und niedrige Raumfrequenzen nicht mehr die gleiche Intensität. Um die Details wie α -Helices und β -Faltblätter in der Struktur aufzulösen, müssen diese Intensitätsunterschiede angepasst werden (Glaeser and Downing, 1992; Conway and Steven, 1999; Rosenthal and Henderson, 2003). Dieses Verfahren nennt sich *b-factor sharpening*.

Um die Raumfrequenzen im richtigen Verhältnis zu verstärken, ist es von Vorteil die Intensität der verschiedenen Raumfrequenzen der jeweiligen Struktur zu kennen, z. B. über ein bekanntes atomares Modell. Dieses Modell wurde mithilfe des stabilen Einpassens (Kapitel B.6.12) in die 3D-Rekonstruktion eingefügt. Das atomare Modell wurde gespeichert und eine neue Dichtekarte mit dem Modul `sxpdb2em.py` simuliert. Durch die in Abb. B.21 gezeigten Befehle innerhalb der Sparx-Kommandozeile wurde eine Frequenzdatei erstellt, die den Faktor der Verstärkung angab (Penczek, 2010a). Die Zeilen 8 bis 12 können in die Anwenderfunktion integriert werden, damit das *sharpening* automatisch nach jeder Iteration ausgeführt wird.

Falls keine atomare Struktur vorhanden ist, gibt es Programme die das Verhältnis der Intensitätskorrektur berechnen (Rosenthal and Henderson, 2003; Fernández et al., 2008). Diese Verfahren kam in dieser Arbeit aber nicht zur Anwendung.

```
from EMAN2_cppwrap import *
2 from global_def import *
from sparx import *
4
def ref_ali3d( ref_data ):
6     from applications import header
    from utilities import print_msg
8     from filter import fit_tanh, filt_tanl
    from fundamentals import fshift
10
    cs = [0.0]*3 # definition des Zentrums
12
    stat = Util.infomask(ref_data[2], ref_data[0], False)
14    volf = ref_data[2] - stat[0]
    Util.mul_scalar(volf, 1.0/stat[1])
16    Util.mul_img(volf, ref_data[0]) # Maskieren der 3D-Rekonstruktion
    fl, aa = fit_tanh(ref_data[3]) # Bestimmen der Filterparameter
18    msg = "Tangent filter: cut-off frequency = %10.3f          fall-off = %10.3f\n"
    n"%(fl, aa)
    print_msg(msg) # Ausgabe der Filterparameter in der Logdatei
20    volf = filt_tanl(volf, fl, aa) # Eigentliches Filterprozess
22
    if ref_data[1] == 1: # Falls die Rekonstruktion zentriert werden soll
        cs = volf.phase_cog() # bestimmen des Zentrums
24        msg = "Center x = %10.3f          Center y = %10.3f          Center z =
%10.3f\n"%(cs[0], cs[1], cs[2])
        print_msg(msg) # Werte für die Verschiebung ausgeben
26        volf = fshift(volf, -cs[0], -cs[1], -cs[2]) # Eigentliche Verschiebung
    return volf, cs # Ergebnis an Sparx zurückgeben
```

Abb. B.22: Anwenderfunktionsdatei für Sparx in der Programmiersprache Python

Filtern mit der Anwenderfunktionsdatei

Um die Ausrichtung an Rauschen zu verhindern, wurde die 3D-Rekonstruktion vor der Erstellung der Referenzen (siehe Abschnitt B.6.8) gefiltert. Dieser Prozess geschah im Fourierraum. Da sich das hochfrequente Rauschen eher am Rand, die niederfrequenten Proteindetails dagegen in der Mitte befanden, konnte das Rauschen durch eine Kreismaske entfernt werden.

Um den Radius dieser Kreismaske entsprechend zu bestimmen verwendete Sparx bevorzugt eine Anwenderfunktion (*userfunction*), die nach jeder Iteration ausgeführt wird. Hier können neben dem Filtern der 3D-Rekonstruktion gleichzeitig der *Sharpening*-Prozess ausgeführt oder beliebige weitere Funktionen eingebaut werden. Abb. B.22 zeigt eine solche Anwenderfunktion.



6.8 Referenzenerstellung

Zur Ausrichtung der Partikel werden Projektionen von der 3D-Rekonstruktion erzeugt, die dann als Referenz dienen. Dafür wird die Rekonstruktion in gleichmäßigem Abstand um die drei Raumwinkel gedreht, bis von allen Orientierungen eine Projektion des Moleküls vorhanden ist. Die Schrittweite in Grad ($^{\circ}$) variierte in der Regel zwischen 15 bis 1° , abhängig von der jeweiligen Auflösung. Durch die kleineren Schritte werden feinere Unterschiede der 3D-Rekonstruktion in den Projektionen abgebildet, allerdings steigt die Rechenzeit durch die höhere Anzahl an Referenzen an. Ein Beispiel für Referenzen zeigt Abb. B.23. Durch kleinere Winkel bei der Referenzenerstellung kann die Ausrichtung verfeinert werden und dadurch sich die Auflösung verbessern.

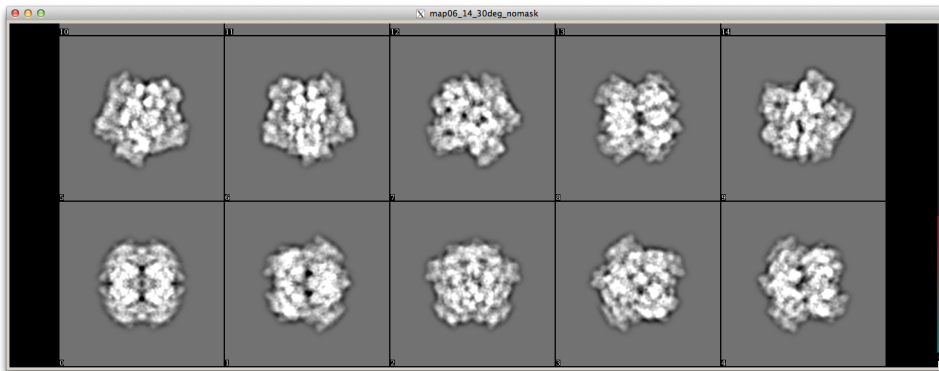


Abb. B.23: Teil der Referenzen von *L. polyphemus* mit gleichmäßigem Abstand von 30° .

6.9 Partikelausrichtung

Allgemein lässt sich die Ausrichtung im 2D-Raum als ein spezielles Problem des Abgleichs mit einer Vorlage (*template matching*) beschreiben. In der Einzelpartikelanalyse werden Referenz als Vorlage benutzt. Es gibt verschiedene Algorithmen, die sich in ihrer Effizienz und Genauigkeit unterscheiden. Die Ausrichtung mit Umwandlung in Polarkoordinaten (*alignment using resampling to polar coordinates*) hat sich als eine der effizientesten und genauesten durchgesetzt (Joyeux and Penczek, 2002).

Bei diesem Algorithmus wird der jeweilige Partikel an einer Referenz ausgerich-

tet. Dazu müssen zwei translationale Parameter, x und y , und ein Rotationswinkel φ bestimmt werden. In Bezug auf die Translation kann es hier zur Interpolation kommen, wenn der Partikel nicht genau um einen Pixel k verschoben wird ($k + x$; $x = 0, 1$). Bei einer Translation von $k + x$; $x =]0, 1[$ Pixel findet eine Interpolation statt und ist am größten bei $x = 0.5$. Bei der Rotation kommt es ebenfalls zu Interpolationsartefakten, wenn $\varphi \neq m \cdot 90$; $m = \{0, 1, 2, 3\}$ entspricht. In den meisten Ausrichtungsmethoden können sich die Artefakte akkumulieren, ausgenommen bei der Umwandlung in Polarkoordinaten. Sowohl die Referenz als auch das auszurichtende Bild werden dabei nur ein mal mit dem gleichen Algorithmus umgewandelt (Joyeux and Penczek, 2002).

Hierfür werden von den auszurichtenden Partikeln mehrere konzentrische Kreise mit definiertem inneren und äußeren Radius erstellt. Dazu wird der Mittelpunkt in x - und y -Richtung verschoben, entsprechend der eingestellten Schrittweite und maximalen Verschiebung. Neben den konzentrischen Kreisen der Partikel wird für die Referenzen jeweils ein konzentrischer Kreis im Mittelpunkt des Partikels erstellt. Diese Kreise werden mithilfe der eindimensionalen (1D) FFT transformiert und per Kreuzkorrelation verglichen. Die höchste Korrelation bestimmt demnach die Verschiebung, die Position des Maximums definiert die Rotation. Diese Werte dienen als neue Grundlage für die folgende 3D-Rekonstruktion.

6.10 Qualität und Auflösung der Rekonstruktion

Die Zwischenergebnisse der Rekonstruktionen sowie das Endergebnis müssen hinsichtlich ihrer Qualität bewertet werden. Dafür hat sich der Begriff der Auflösung durchgesetzt, wobei in diesem Fall die Auflösung der 3D-Rekonstruktion mit der optischen Auflösung eines Mikroskops nicht zu vergleichen ist (Penczek, 2010b) (siehe Kapitel B.3.1).

In der Einzelpartikelanalyse wird die Auflösung durch die Korrelation der verschiedenen Raumfrequenzen bestimmt. Hierzu wird der Partikelstapel zufällig in zwei Teile geteilt und getrennt 3D-rekonstruiert. Durch die Transformation in den Fourierraum kann man Schalen einer spezifischen Frequenz miteinander korrelieren. Die Werte werden in der Fourier-Schalen-Korrelation (*fourier shell correlation*) (FSC) aufgetragen (Abb. B.24) (Harauz and van Heel, 1986; Orlova et al., 1997).

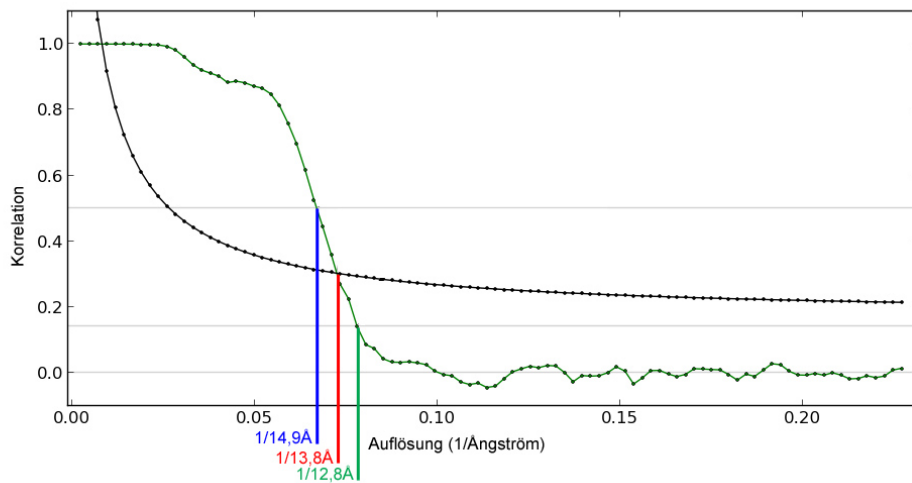
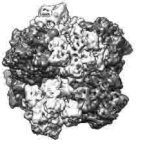


Abb. B.24: Bei der FSC wird die Korrelation der Raumfrequenzen zweier 3D-Rekonstruktion in einem Graphen aufgetragen. Der Grenzwert zum Ablesen der Auflösung ist abhängig von der verwendeten Methode zur Rekonstruktion. Häufig verwendet wird der Punkt, an dem die Korrelation unter 50 % fällt (0.5-Kriterium) (blau), das 1/2 bit-Kriterium (rot) und das 0.143-Kriterium (grün).

Die Wahl der Messpunkte ist abhängig von der Rekonstruktionsmethode. Generell wird der x -Wert am Schnittpunkt mit der 1/2 bit-Kurve (van Heel and Schatz, 2005) bzw. des jeweiligen y -Wertes der Korrelation abgelesen. Dabei ist das 0.5-Kriterium (Beckmann, 1997; Böttcher et al., 1997; Conway et al., 1997; van Heel and Schatz, 2005) das restriktivste, das 0.143-Kriterium (Rosenthal and Henderson, 2003) das am wenigsten restriktive. Das 0.143-Kriterium kam zum Einsatz, wenn mit zwei komplett unabhängigen Datensätzen gearbeitet wurde. In Abb. B.24 werden alle Kriterien angezeigt, die richtige Auswahl bleibt dem Benutzer überlassen.

Abgesehen von der oben genannten Korrelationsmessung, kann die Auflösung auch visuell bestimmt werden. Ist eine bestimmte Auflösung erreicht (kleiner 8 \AA), kann man in der 3D-Struktur die ersten Sekundärstrukturmerkmale wie α -Helices erkennen. Ab einer Auflösung von etwa $4,5 \text{ \AA}$ ist das $C\alpha$ -Rückrad der Aminosäuresequenz und β -Faltplättler zu sehen. Bei Auflösungen kleiner 4 \AA werden dann Ausbuchtungen in der Dichtekarte sichtbar, die auf aromatische Aminosäuren schließen lassen (Zhou, 2008).

6.11 Iterative Verfeinerung

Nach der Erzeugung der ersten 3D-Rekonstruktion wurden die Schritte in Kapitel B.6.6 bis B.6.10 iterativ in einer Schleife ausgeführt. Die Winkel zur Erzeugung der Referenzen wurden schrittweise immer kleiner gewählt. Abhängig von der Auflösung, wurden zusätzliche Methoden hinzugefügt, die die Qualität der Rekonstruktion weiter verbessern konnten.

Im Ergebnisteil von *Limulus polyphemus* (Kapitel C.3.1 ab Seite 64) wurden die verwendeten Parameter und Zeitpunkte während des Verfeinerungsprozess genau aufgeführt und sollen hier als Beispiel dienen.

6.12 Einpassen atomarer Modelle

Beim Einpassen atomarer Modelle in die Elektronendichtekarte gibt es zwei unterschiedliche Ansätze. Das atomare Modell, z. B. aus der Röntgenkristallographie, wird in die Dichtekarte eingepasst ohne die Tertiärstruktur zu verändern (*rigid body fitting*). Dabei wird es translational und rotational verschoben, bis die Korrelation mit der Dichtekarte maximal ist. Die zweite Möglichkeit ist das flexible *Fitten* (Einpassen). Dabei darf sich die Konformation des atomaren Modells in bestimmten Grenzen bewegen und die Korrelation mit der Dichtekarte kann weiter maximiert werden.

In der Regel werden beide Techniken in Kombination benutzt. Für das starre Einpassen kann das Modul `Fit in Map` (Goddard et al., 2007) aus dem Programm `Chimera` verwendet werden. Falls die Auflösung der 3D-Rekonstruktion 15 Å und besser (Schröder et al., 2007) erreicht hat, kann das flexible *Fitting* beispielsweise mit der Software `Direx` im Anschluss durchgeführt werden.

Der Algorithmus von `Direx` simuliert schrittweise eine Dichtekarte der atomaren Struktur und bestimmt die Korrelation mit der EM-Dichtekarte. Danach werden die Atome, die sich in einer geringeren Dichte befinden, in Richtung der höheren Elektronendichte bewegt (Wang and Schröder, 2012) und eine neue Dichtekarte wird simuliert. Falls die Korrelation höher war, wird mit der neuen Struktur weiter gearbeitet. Das Programm benutzt dabei genetische Algorithmen, um den Grad der Veränderungen der atomaren Struktur zu steuern. Parameter prüfen die Plausibilität der Veränderungen (de Groot et al., 1997; Schröder et al., 2007; Wang and Schröder, 2012).



6.13 Molekulares Modellieren mit Gromacs

Mit **Gromacs** werden Moleküldynamik-Simulationen (MD-Simulationen) berechnet. Dabei wird die Bewegung und Geschwindigkeit einzelner Atome im Gesamtsystem modelliert. Dabei wird das zu simulierende Molekül in einen Raum mit periodischen Randbedingungen integriert. Moleküle sowie Kräfte können durch die Ränder auf der gegenüberliegenden Seite wieder eintreten. Wichtig bei der Berechnung der Kräfte ist die Wahl des Kraftfeldes. Ist das verwendete Modell richtig gewählt, ist es mit Moleküldynamik (MD)-Simulationen möglich exakte Ergebnisse mit der Limitierung der numerischen Präzision zu erreichen (Sutmann, 2002). Das Ergebnis stellt eine Trajektorie für jedes einzelne am System beteiligte Atom dar, mit dem die Bewegung im Raum verfolgt werden kann.

Voraussetzung für exakte Ergebnisse sind die richtigen Umgebungsbedingungen wie Pufferzusammensetzung und die Anfangskonfiguration. Ist diese weit entfernt vom energetischen Gleichgewicht, kann die Moleküldynamik (*molecular dynamics*) (MD)-Simulation fehlschlagen. Um dies zu vermeiden, ist eine Energieminimierung essenziell (Hess et al., 2013). Dadurch lässt sich auch das thermische Rauschen am Anfang der Simulation minimieren, das bei der gemeinsamen Auswertung mehrerer Simulationen besonders wichtig ist.

Zum Ergebnis einer MD-Simulation mit **Gromacs** ist zu sagen, dass nach dem Start der Simulation immer nur das nächste lokale Minimum gefunden wird. Um repräsentative Aussagen treffen zu können, müssen mehrere Simulationen vom gleichen Startzustand erstellt werden.

6.14 Verwendete Module

Eine Liste der benutzten Module mit dem zugehörigen Programm sowie eine kurze Beschreibung, wird in Tabelle B.4 aufgeführt.

Tab. B.4: Die verwendeten Programmmodule, das zugehörige Programm und die Funktion

Modulname	Software	Funktion
sxcpy.py	Sparx	.hdf Dateien kopieren und aneinanderfügen
sxisac.py	Sparx	Reproduzierbare Klassensummen erstellen
sxali2d.py	Sparx	2D Ausrichtung von Partikeln mithilfe einer Referenzfreien Ausrichtung
sxali3d.py	Sparx	3D Ausrichtung und Rekonstruktion anhand eines Referenzvolumens
sxheader.py	Sparx	Header von Dateien auslesen Bilderstapel auslesen
sxk_means.py	Sparx	Klassensummen erstellen
sxmref_ali2d.py	Sparx	2D Ausrichtung anhand von Referenzen
sxproject3d.py	Sparx	2D Projektionen eines 3D Volumens erzeugen
sxpdb2em.py	Sparx	Simulieren einer Elektronendichtekarte anhand einer PDB Datei
sxfind_struct.py	Sparx	Finden der Eulerwinkel einer von mehreren 2D Projektionen
sxtransform2d.py	Sparx	2D Alignment auf die Partikel anwenden
e2proc2d.py	EMAN2	2D Bilderstapel verarbeiten (aufteilen, verkleinern, etc.)
e2proc3d.py	EMAN2	3D Volumen modifizieren (Filtern, Maskieren, etc.)
e2display.py	EMAN2	Zum anzeigen von Bilderstapel
e2iminfo.py	EMAN2	Anzeigen von Bildinformationen
direx	Direx	Zum flexiblen Fitten von Kristall-Strukturen in Elektronendichtekarten
apply-cos-mapping	Direx	Zum bearbeiten der Maske für DireX
pdb2gmx	Gromacs	Wandelt eine PDB in ein passende Format für Gromacs um
editconf	Gromacs	Erstellt eine Box um ein Molekül mit spezifischen Abstand
genion	Gromacs	Verteilt Ionen innerhalb der erstellten Box
grompp	Gromacs	Erzeugt eine Datei für die Molekulare Dynamik Simulation
mdrun	Gromacs	Startet die Molekulare Dynamik simulation
g_dist	Gromacs	Messung von Atomabständen in einer Trajektorien-datei
ctffind3.exe	ctffind3	Zum finden der Defokuswerte von elektronmikroskopischen Aufnahmen
boxer	EMAN190	Zum manuellen selektieren von Partikeln
proc3d	EMAN190	3D Volumen modifizieren (maskieren, filtern, etc.)
refine	EMAN190	3D Ausrichtung und Rekonstruktion anhand eines Referenzvolumens
measure rotation	Chimera	Bestimmt den Grad der Rotationssymmetrie
FindHBond	Chimera	Wasserstoffbrücken identifizieren
Fit in Map	Chimera	Einpassen von Strukturen in Dichtekarten
MatchMaker	Chimera	Erstellung eines Homologiemodells
Manual Picker	Appion	Manuelles selektieren von Proteinen auf EM-Bildern



7 Programmiersprachen und -methoden

7.1 Programmiersprachen

Python

Der Großteil der Programmierarbeiten für diese Arbeit wurde in der Skriptsprache Python entwickelt (van Rossum and Drake, 2002). Python ist eine freie, interpretierte Programmiersprache, die bereits viele Bibliotheken für wissenschaftliche Aufgabenstellungen integriert hat. Es bildet dabei eine ideale Plattform, um moderne Softwarepakete zu entwickeln. Da diese unter ständiger Weiterentwicklung stehen, ist es besonders einfach den Programmcode zu schreiben, zu testen und einer breiten Gruppe an Wissenschaftlern zur Verfügung zu stellen (Hohn et al., 2007).

Es wurden folgende Bibliotheken benutzt:

- PyQt
- matplotlib
- libtif
- numpy
- argparse
- MySQLdb

AWK³

AWK (Aho et al., 1979) ist eine freie Skriptsprache um einfache Textdokumente auszuwerten oder zu manipulieren (Herold, 1991). Sie eignet sich besonders gut für CSV-Dateien (kommaseparierte Werte *comma-separated values*). Es handelt sich um ein einfaches Datenformat, in dem Informationen durch Kommata (aber auch Tabulatoren oder Leerzeichen) getrennt sind. Dieses einfache Format wird in der Wissenschaft sehr häufig verwendet, z. B. in der Einzelpartikelanalyse, um die Winkelinformationen der jeweiligen Partikel zu speichern.

³die Bezeichnung entstand aus den Anfangsbuchstaben der Entwickler Aho, Weinberger und Kernighan

Bash Shell Skript

Die auf dem Computercluster benutzte Kommandozeile war die *Bourne-again shell* (`bash`) (Herold, 1999). Bei der `bash` handelt es sich um eine Weiterentwicklung der Bourne-Shell, sie stellte die Oberfläche für alle Skripte und kommandozeilenbasierten Programme dar.

7.2 Test-driven Development

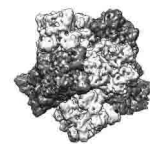
Testgetriebene Entwicklung (*test-driven development*) (TDD) ist eine Programmieretechnik, die einen Benutzer dazu auffordert, jeden Teilschritt seines Programms mit automatischen Tests zu überprüfen. Der Vorteil spiegelt sich in Wiederverwertbarkeit und Lesbarkeit des Programmcodes wider (Beck, 2002). Durch die automatischen Tests des Programms ist auch das Editieren und Erweitern für unerfahrene Programmierer sicher, da jederzeit geprüft werden kann ob die Kernfunktionen des Programms weiterhin gegeben sind (Freeman and Pryce, 2009; Ziade, 2008). Im Fehlerfall wird das Programm nicht in den Regelbetrieb eingeführt und die Fehler müssen erst behoben werden.

Bei TDD wird der Entwicklungsablauf um das Erstellen und Durchführen der Tests erweitert, bevor der zu entwickelnde Programmcode geschrieben wird. Beim Durchführen des Tests wird dieser in erster Instanz fehlschlagen, da er vor dem eigentlich Programmcode geschrieben wurde. Danach kann der entsprechende Quelltext angepasst werden, damit die Bedingung des Tests erfüllt ist. Nachdem man einen automatischen Test für eine programmierte Funktion hat, wird der Quelltext refaktoriert⁴.

7.3 Clean-Code

Bei der Entwicklung von Programmcode wird versucht durch ausführliche Bezeichnung von Variablennamen und Funktionsnamen auf Kommentare im Quelltext komplett zu verzichten (Martin, 2008), da diese beim Refaktorisieren oft nicht übernommen werden und der Quelltext dadurch schwerer lesbar wird. Weiterhin werden klare Regeln für den Namensraum im Programm gegeben, der dabei hilft den Quelltext zu vereinheitlichen.

⁴Manuelle oder automatisierte Strukturverbesserung von Quelltexten



C Ergebnisse

1 3D-EM des multimeren Hämoglobins von *Daphnia pulex*

Bei den respiratorischen Proteinen von Wasserflöhen der Gattung *Daphnie* handelt es sich um extrazelluläre multimere Hämoglobine, die sich frei in der Hämolymphe der Tiere befinden. Die Hämoglobine von *Daphnia pulex* und *Daphnia magna* wurden von [REDACTED] aufgereinigt und mir zur Verfügung gestellt. In der Literatur bestand Uneinigkeit hinsichtlich der Anzahl der Häm-Domänen in diesen beiden Hämoglobinen. Die Angaben schwankten zwischen 24 und 28 bei *D. pulex* und 32 bei *D. magna* (Peeters et al., 1990; Dewilde et al., 1999; Lamkemeyer et al., 2005; Dangott and Terwilliger, 1980). Ein Vergleich im Elektronenmikroskop zeigte jedoch deutlich eine Ähnlichkeit beider Hämoglobine sowohl in der Größe als auch in der Form (Abb. C.1). Daraufhin wurde nur noch mit dem Hämoglobin von *D. pulex* weitergearbeitet.

Für die Analyse mittels 3D-EM des Hämoglobins von *D. pulex* wurden zwei Datensätze erstellt. Ein Datensatz mit Negativkontrast-EM-Aufnahmen wurde mithilfe einer Digitalkamera aufgenommen und diente der Erzeugung eines Startmodells. Der Zweite Datensatz wurde von kryogenisch präparierten Proben auf Negativ-Filmen erstellt und anschließend mit einem Trommelscanner (siehe B.3.4) bei einer Auflösung von 0,985 Å/Pixel digitalisiert. Um hohe Rechenzeiten zu vermeiden, wurde die Scanauflösung halbiert und mit einer Pixelgröße von 1,97 Å/Pixel gearbeitet. Zur manuellen Selektion der Partikel wurde in beiden Fällen das EMAN-Modul `boxer` benutzt. Die vollständigen technischen Daten sind in Tabelle C.1 aufgeführt.

Der Datensatz aus negativkontrastierten EM-Bildern wurde mit den Sparx-Modulen `sxisac`, `sxfind_struct` und `sxali3d` bearbeitet und ein Startmodell

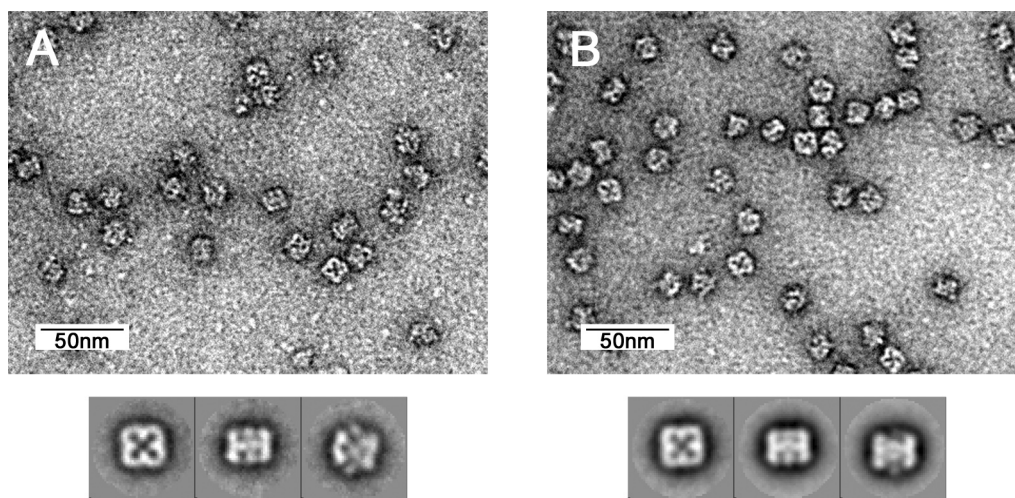


Abb. C.1: Elektronenmikroskopische Aufnahmen und die häufigsten Klassen-summen negativkontrastierten *Daphnia*-Hämoglobins. (A) Hämoglo-bin von *D. pulex*. (B) Hämoglobin von *D. magna*. Weder in der Größe noch in der Form der Moleküle ist ein signifikanter Unterschied zu erkennen.

Tab. C.1: EM-Datensätze des *Daphnia pulex*-Hämoglobins

	Kryo-EM-Aufnahmen	Negativkontrast-EM-Aufnahmen
Anzahl der Partikel:	31 481	6 480
Anzahl der EM-Bilder:	74	51
Boxgröße:	144 × 144	50 × 50
Finale Auflösung:	14 Å nach 0.5-Kriterium	nicht bestimmt
Vergrößerung:	62 950×	nominell 82 000×
Ångstrom pro Pixel:	1,97	ca. 6,4
Defokus:	−1,2 μm bis −5,5 μm	−1 μm bis −5 μm
Elektronenmikroskop:	Tecnai G ² F20	Tecnai G ² 12 BioTWIN

des Hämoglobins erstellt. Hier wurde keine weitere Auflösungsbestimmung vor-genommen, da es im Anschluss als Referenz für die weitere Verfeinerung am Datensatz aus Kryo-EM-Bildern benutzt wurde. Hier konnte ich schon nach weni-gen Iterationen durch den Befehl `measure rotation` im Programm `Chimera` eine D4 Symmetrie bestimmen. Danach wurde der Datensatz mit dem `Sparx`-Modul `sxali3d` verfeinert, bis keine Verbesserung der Auflösung mehr zu messen war. Um die Prozessierung von Rauschen zu vermeiden wurden die Filterwerte auto-

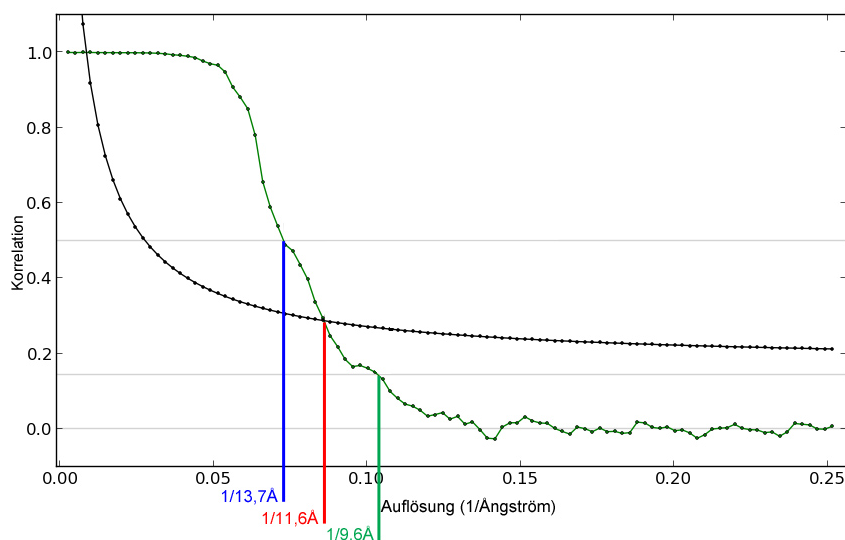


Abb. C.2: Fourier-Schalen-Korrelation (*fourier shell correlation*) der finalen 3D-Rekonstruktion des Hämoglobins von *D. pulex*. Aufgrund des verwendeten Verfahrens zur Rekonstruktion wurde das 0.5-Kriterium (blau) als relevant angenommen.

matisch 2 \AA höher gewählt, als die jeweilige FSC berichtete. Die finale Auflösung der 3D-Rekonstruktion betrug ca. 14 \AA nach dem 0.5-Kriterium (Abb. C.2).

Da aufgrund der hier ermittelten D4-Symmetrie des Hämoglobins die Anzahl der Untereinheiten durch 4 teilbar sein muss, kamen für die Quartärstruktur nur 12 oder 16 Untereinheiten in Frage. 16 Untereinheiten entspricht einer Masse von ca. 560 kDa. Mit entsprechend eingestelltem Schwellenwert ergaben sich für das Hämoglobin eine Höhe von ca. 110 \AA und eine Breite von ca. 140 \AA (Abb. C.3 A, B).

Da keine Kristallstruktur der dimeren Untereinheit vorlag, wurden mithilfe des Chimera-Moduls `molmap` einzelne Globine mit einer Auflösung von 14 \AA simuliert und in die finale Rekonstruktion eingepasst. Durch dieses Verfahren war es möglich abzuschätzen, wieviele Globine gleichmäßig in das Molekül passten (Abb. C.3 C, D). In Bezug auf die D4-Symmetrie ließen sich vier Globine (zwei identische Dimere) sehr leicht in eine asymmetrische Untereinheit einpassen, was auf insgesamt 16 Dimere hin deutet.

Ein auf 420 kDa eingestellter Schwellenwert und nur 12 Dimere stellte sich

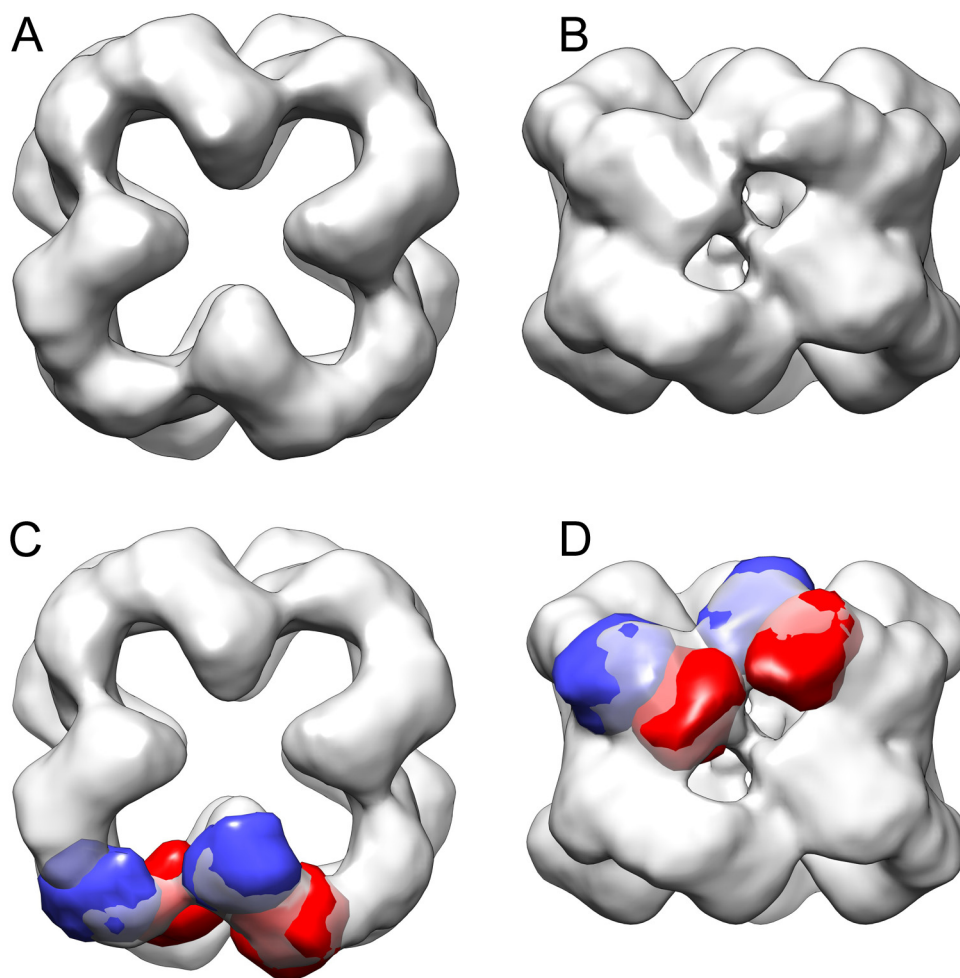


Abb. C.3: 3D-Rekonstruktion von *D. pulex*-Hämoglobin bei 560 kDa massenkorreliertem Schwellenwert. (A) Aufsicht entlang der D4-Symmetrie-Achse mit einer Länge und Breite von ca. 140 Å. (B) Seitenansicht mit einer Höhe von ca. 110 Å. (C), (D) Passform von vier auf 14 Å simulierten Häm-Domänen in eine der acht asymmetrischen Einheiten des Moleküls. Die beiden so entstandenen Dimere (rot/blau) sind identisch. Es bleiben keine großen Massen leer.

als weniger vorteilhaft heraus, denn größere Bereiche der Dichtekarte wurden von den Dimer-Modellen nicht ausgefüllt, wohingegen andere die Modelle nicht vollständig umschlossen (Abb. C.4). Zudem müssten für eine solche Struktur zwei unterschiedliche Typen von Dimeren vorhanden sein.

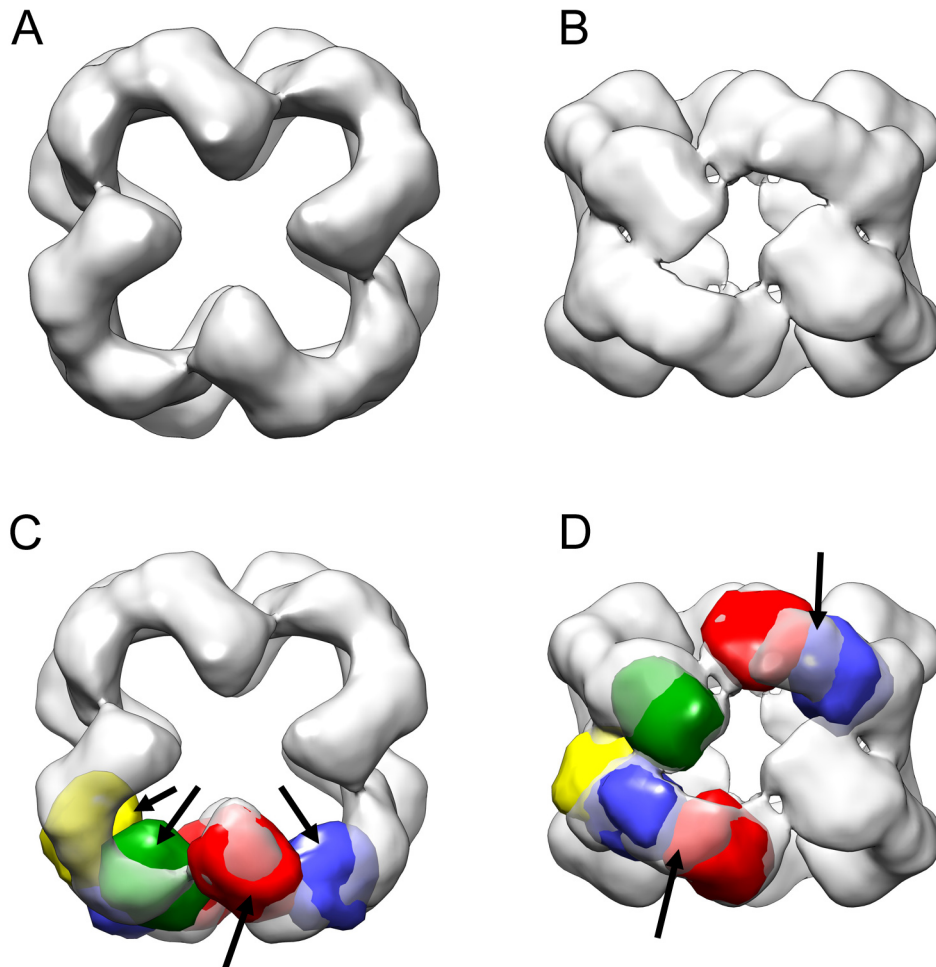
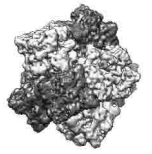


Abb. C.4: 3D-Rekonstruktion von *D. pulex* Hämoglobin bei 420 kDa massenkorreliertem Schwellenwert. A) zeigt das Molekül entlang der D4-Symmetrie-Achse. B) zeigt die Seitenansicht des Moleküls, C) und D) Passform von vier auf 14 Å simulierten Häm-Domänen in zwei der acht asymmetrischen Einheiten des Moleküls. Die zwei Dimere (rot/blau) sind identisch, das dritte Dimer (grün/gelb) stellt einen weiteren Dimer-Typ dar. (C) Die Pfeile weisen auf zu wenig Masse für die Häm-Domänen hin. (D) Die Pfeile zeigen auf leere Bereiche in der 3D-Rekonstruktion.

2 3D-EM des 2x6meren Hämocyanins von *Astacus leptodactylus*

Das Hämocyanin von *A. leptodactylus* ist ein 2x6mer (Abb. C.5), wobei die Kontaktstelle der Hexamere zueinander nicht genau bekannt war. Hier sollte eine Disulfidbrücke existieren (Jeffrey et al., 1978; Markl and Decker, 1992)

Die genaue Anordnung der beiden Hexamere zueinander sollte mithilfe der 3D-EM untersucht werden. Hierzu wurde das Hämocyanin aufgereinigt (siehe B.2.1) und ein Kryo-EM-Datensatz mit 11 832 Einzelpartikeln erstellt (Tabelle C.2).

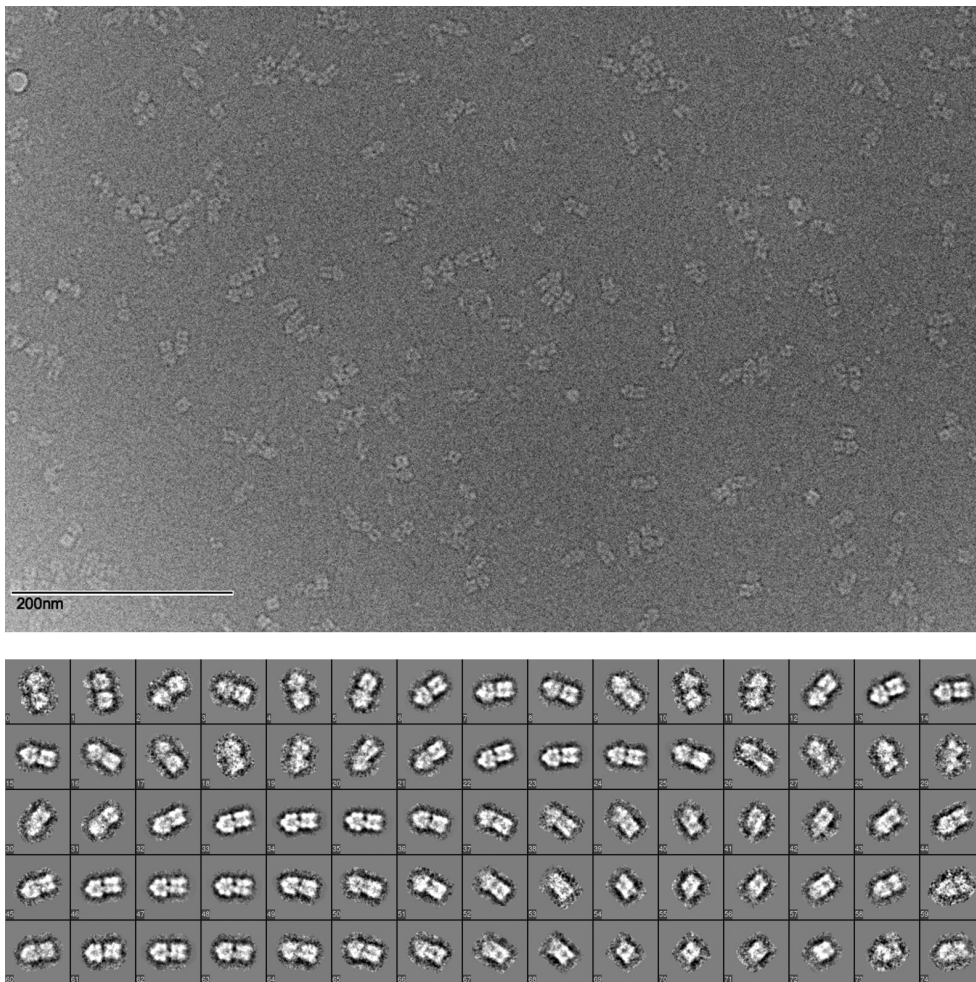
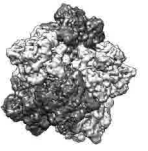


Abb. C.5: Kryo-EM-Aufnahme von *Astacus leptodactylus*-Hämocyanin (oben), und ein Teil der Klassensummen aus dem Datensatz.



3D-EM des 2x6meren Hämocyanins von *Astacus leptodactylus*

Tab. C.2: Kryo-EM-Datensatz des *Astacus leptodactylus*-Hämocyanins

Anzahl der Partikel:	11 832
Anzahl der EM-Bilder:	142
Boxgröße:	192 × 192
Finale Auflösung:	15 Å nach 0.5-Kriterium
Vergrößerung:	71 540×
Ångstrom pro Pixel:	2,18
Defokus:	−4,99 µm bis −1,25 µm
Elektronenmikroskop:	Tecnai G ² 12 BioTWIN

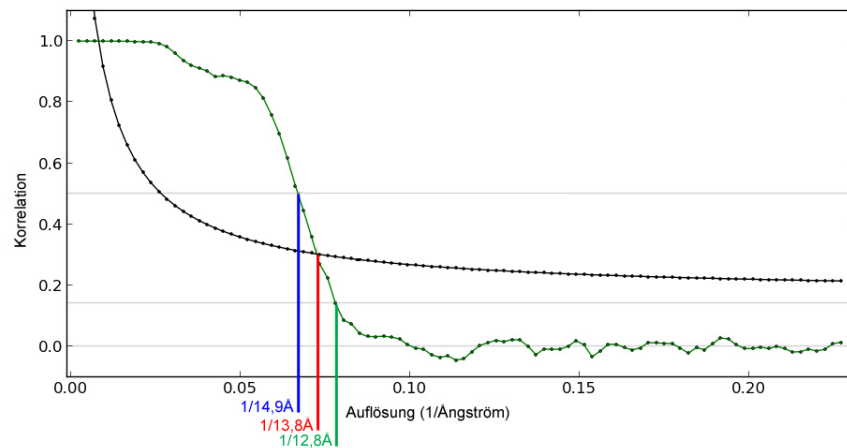


Abb. C.6: FSC der letzten Rekonstruktion von *A. leptodactylus*. Aufgrund des verwendeten Verfahrens zur Rekonstruktion war an dieser Stelle das 0.5-Kriterium (blau) mit 14,93 Å zu berücksichtigen.

Die Partikel wurden mithilfe des Appion-Moduls `Manual Picker` manuell selektiert. Als Startmodell diente eine ältere, auf 30 Å gefilterte 3D-Rekonstruktion aus Negativkontrast-EM-Aufnahmen (zur Verfügung gestellt von [REDACTED]). Das Startmodell wurde mit dem Sparx-Modul `sxali3d` an den Kryo-Partikeln verfeinert. Das Ergebnis war eine 3D-Rekonstruktion mit 15 Å Auflösung nach dem 0.5 Kriterium (Abb. C.6). Das rekonstruierte 2x6mer war bei massenkoreliertem Schwellenwert (900 kDa) 240 Å lang und 145 Å breit. Die Zentren der beiden Hexamere waren 110 Å voneinander entfernt und um 15 Å gegeneinander verschoben (Abb. C.7).

Durch Einpassen des kristallographischen 1x6mers des *Panulirus interruptus*-

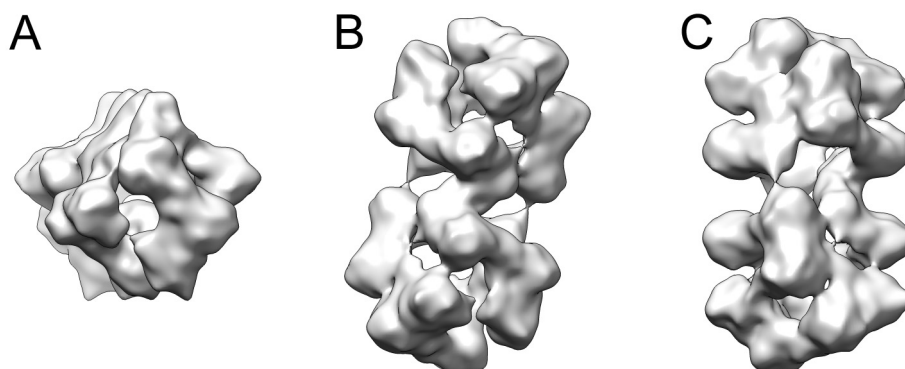


Abb. C.7: 3D-Rekonstruktion des 2x6meren Hämocyanins aus *Astacus leptodactylus*. (A) Blick entlang der y -Achse. (B) Blick entlang der z -Achse (Hauptsymmetrieachse). (C) Blick nach Drehung um 90° um die y -Achse.

Hämocyanins (PDBid: 1HCY (Volbeda and Hol, 1989)) mit dem **Chimera**-Modul **Fit in Map**, ließ sich der Rotationswinkel der Hexamere zueinander genau bestimmen. Dabei waren die Rotationssymmetrieachsen (rot in Abb. C.8) um 103° gegen den Uhrzeigersinn gegeneinander verdreht, die Achsen standen dabei fast parallel (ca. 3° Kippung). In der Rekonstruktion zeigte sich bei massenkorelliertem Schwellenwert eine deutliche Verbindung zwischen den beiden Hexameren. Im Bereich dieser Brücke lag jeweils das 7-strängige β -Barrel (Domäne 3) der betroffenen Untereinheit. Die Aminosäuren H420 näherten sich der jeweiligen Untereinheit des anderen Hexamers bis auf etwa 4 \AA (Abb. C.9).

An anderen Positionen waren dünnere Brücken zwischen den Hexameren zu sehen, bei denen es sich jedoch um Artefakte handeln könnte, denn sie verschwanden beim Reduzieren des Schwellenwertes sehr schnell. Aufgrund der Auflösung der 3D-Rekonstruktion von nur 15 \AA verzichtete ich auf eine automatisches Einpassen der individuellen Untereinheiten, sondern beschränkte mich darauf, das kristallographische Hexamer als starren Körper einzupassen.

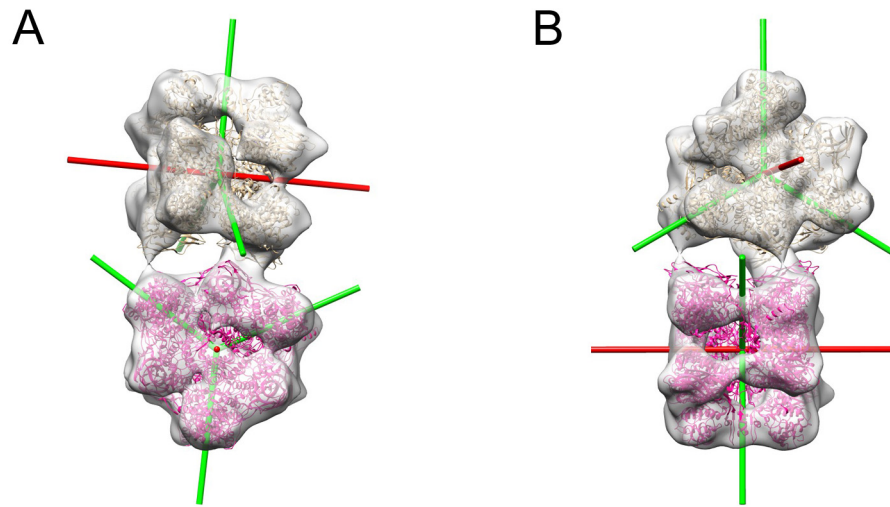
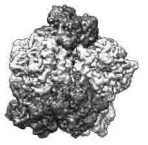


Abb. C.8: Die dreizahlige Symmetrieachsen (rot) der in das *A. leptodactylus*-Hämocyanin eingepassten Hexamere waren um 103° gegen den Uhrzeigersinn gegeneinander verdreht. Dabei war sie um 3° gekippt. An der breiten Kontaktstelle der beiden Hexamere lag jeweils die Domäne 3 der betreffenden Untereinheiten.

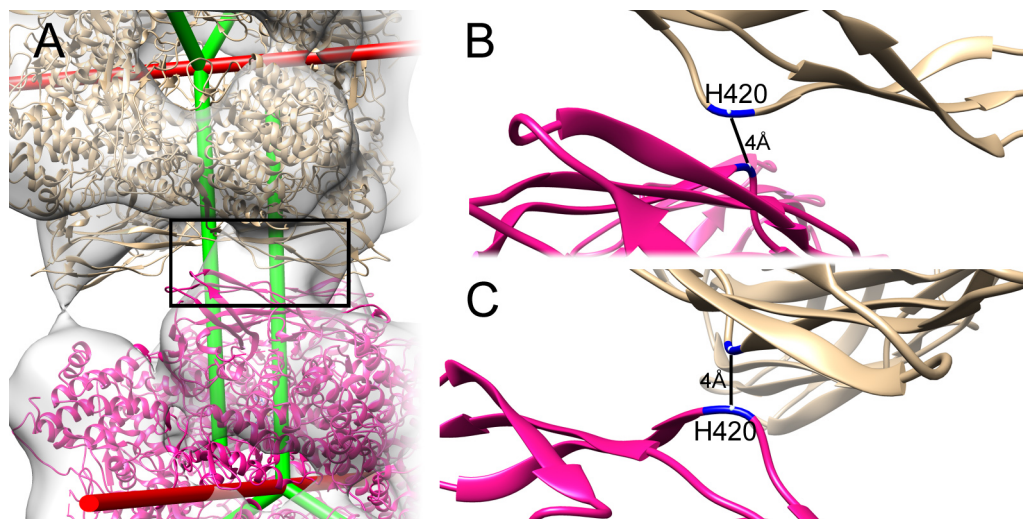


Abb. C.9: Breite Kontaktstelle der beiden Hexamere des *A. leptodactylus*-Hämocyanins. (A) zeigt in dem schwarz umrandeten Kasten die Kontaktstelle. (B) und (C) zeigen die Kontaktstelle um jeweils 90° gedreht mit blau markierter Aminosäure H420 im Abstand von 4 \AA .

3 3D-EM des 8x6meren Hämocyanins von *Limulus polyphemus*

Das Atmungsprotein von *L. polyphemus* ist ein 8x6meres Hämocyanin, dessen 3D-Struktur als ca. 12 Å-Rekonstruktion und molekulares Modell bekannt war (Martin et al., 2007) und nun mittels der 3D-EM genauer analysiert werden sollte. Hier sollte mit einem von [REDACTED] [REDACTED] zur Verfügung gestellten Kryo-EM-Datensatz und dem Programm Sparx eine hochaufgelöste 3D-Rekonstruktion erzeugt werden, die dann mithilfe des flexiblen Einpassens von Homologiemodellen eine quasi atomare Auflösung erreichen sollte. Eine der Kontaktstellen zwischen den Hexameren sollte danach exemplarisch mithilfe Molekulardynamik-Simulation (MD-Simulation) weiter analysiert werden.

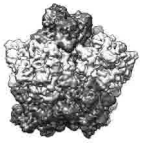
3.1 Erstellen einer 7,1 Å-Rekonstruktion

Der mir zur Verfügung gestellte, am Max-Planck-Institut (MPI) für Biophysik in Frankfurt (AG [REDACTED] [REDACTED]) aufgenommene Kryo-EM-Datensatz des *L. polyphemus*-Hämocyanins (Tabelle C.3) wurde durch Neukalibrierung der Vergrößerung des Elektronenmikroskops neu bearbeitet. Die gescannten Bilder besaßen demnach eine Auflösung von nur 0,96 anstelle von 1 Å/Pixel wie zuvor fälschlich angenommen. Aus diesem Grund wurde die CTF-Korrektur mit entsprechenden Werten erneut durchgeführt. Dann wurde die Auflösung der EM-Bilder auf 1,92 Å/Pixel reduziert.

Zur Erstellung der 3D-Rekonstruktion wurde nach dem Goldstandard-FSC-

Tab. C.3: Kryo-EM-Datensatz des *Limulus polyphemus*-Hämocyanins

Anzahl der Partikel:	38 032
Anzahl der EM-Bilder:	428
Boxgröße:	256 × 256
Finale Auflösung:	7,1 Å nach Goldstandard FSC
Vergrößerung:	61 400×
Ångstrom pro Pixel:	1,92
Defokus:	−5,6 µm bis −0,8 µm
Elektronenmikroskop:	Tecnai G ² Polara



3D-EM des 8x6meren Hämocyanins von *Limulus polyphemus*

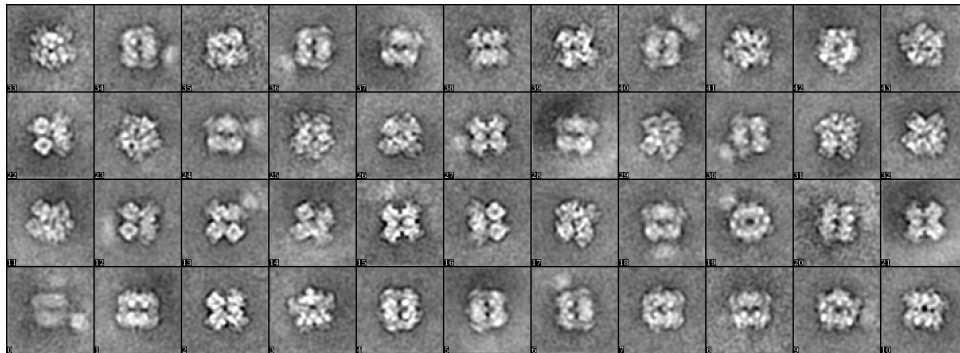


Abb. C.10: Klassensummen des EVEN Datensatzes zur Erzeugung des Startmodells von *Limulus polyphemus*. Der Ausschnitt zeigt 44 Klassen, in die Berechnung des Startmodells wurden ca. 80 Klassen einbezogen.

Verfahren (Rosenthal and Henderson, 2003) gerechnet und somit der gesamte Datensatz in zwei gleich große (19016 Partikel) Sätze geteilt (ODD und EVEN) und identisch behandelt. Um die Klassensummen für das initiale Modell zu berechnen, wurde die Pixelgröße mithilfe des Moduls `e2proc2d.py` und dem Parameter `-meanshrink=4` auf $7,68 \text{ \AA}/\text{Pixel}$ angepasst. Mit `sxisac.py` wurden jeweils ca. 80 Klassensummen (siehe Abb. C.10) erzeugt und die entsprechenden Raumwinkel mit `sxfind_struct.py` bestimmt. Das generierte Startmodell wurde dann für 10 Iterationen mit `sxali3d.py` an den Klassensummen verfeinert.

Unter Verwendung von `e2proc3d.py` wurde das Startmodell auf die Ausgangsgröße von $1,92 \text{ \AA}/\text{Pixel}$ skaliert, und diente als Referenz für die CTF-korrigierten Teildatensätze. Die Winkel zur Referenzerzeugung wurden in 2° -Schritten von 16° auf 8° reduziert und jeweils mit 10 Iterationen verfeinert. Hier kam noch keine Symmetrisierung zum Einsatz. Die Symmetrie wurde an dieser Stelle mit dem Modul `measure rotation (Chimera)` berechnet und bestätigte die früher publizierte D2 Symmetrie des Moleküls (Taveau et al., 1997; Martin et al., 2007). In den nachfolgenden Iterationen wurde diese Symmetrie angewendet. Die Referenzwinkel wurden in 1° -Schritten mit jeweils ca. 15 Iterationen bis auf 3° reduziert. Bei der Verfeinerung kam der Parameter `-stopnrct=0.1` zum Einsatz, der die Iteration abbricht, wenn 99,9% der Partikel eine stabile Ausrichtung eingenommen haben. Somit erfolgte ein automatischer Sprung zum nächst kleineren Winkel.

Nach diesem Durchgang lag die Auflösung bei ca. $9,3 \text{ \AA}$ nach Goldstandard und es konnte der *Sharpening*-Prozess mit in die Verfeinerung aufgenommen werden.

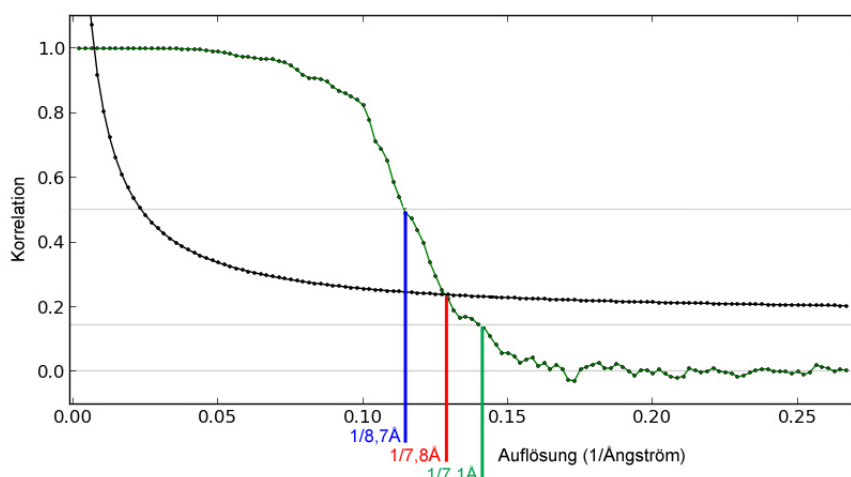


Abb. C.11: FSC der letzten Rekonstruktion von *L. polyphemus*. Aufgrund des verwendeten Verfahrens zur Rekonstruktion ist an dieser Stelle das 0.143-Kriterium (grün) mit 7,1 Å zu berücksichtigen. (Rosenthal and Henderson, 2003)

Dafür wurden die einzelnen Hämocyanin Untereinheiten des bekannten pseudoatomaren Modells von Martin et al. (2007) in die neue Rekonstruktion eingepasst. Anschließend wurde eine Frequenzdatei erstellt und in die Anwenderfunktionsdatei integriert. Die Winkel der Referenzen wurden in 0,5°-Schritten bis auf 1° reduziert. Es ergaben sich 7,9 Å nach Goldstandard.

An dieser Stelle wurden die aktuellen Raumwinkel der einzelnen Partikel auf den unkorrigierten Datensatz übertragen. Daraufhin wurde innerhalb der Verfeinerung der `-CTF` Parameter aktiviert, um `Sparx` anzuzeigen, nun die CTF-Korrektur innerhalb des Algorithmus auszuführen. Zusätzlich wurde mit dem Parameter `-an=4` eine lokale Ausrichtung mit $\pm 4^\circ$ vorgenommen. Damit wurde eine letzte Iteration gerechnet, bis sich 99,9% der Raumwinkel nicht mehr änderten. Die Auflösung betrug nun 7,1 Å nach Goldstandard (Abb. C.11). Gefiltert wurde während der Verfeinerung immer auf den Wert der aktuellen Auflösung +1 Å, um die Ausrichtung an Rauschen zu vermeiden. Das Ergebnis der 7,1 Å-Rekonstruktion ist in Abb. C.12 zu sehen.

Während der gesamten Verfeinerung wurde gemäß Penczek et al. (1992) für die 2D-Ausrichtung der innere Radius auf 5 Pixel gesetzt (`sxali3d.py`-Parameter `-ir`). Hierbei ergab sich im Bereich des von Martin et al. (2007) definierte Inter-

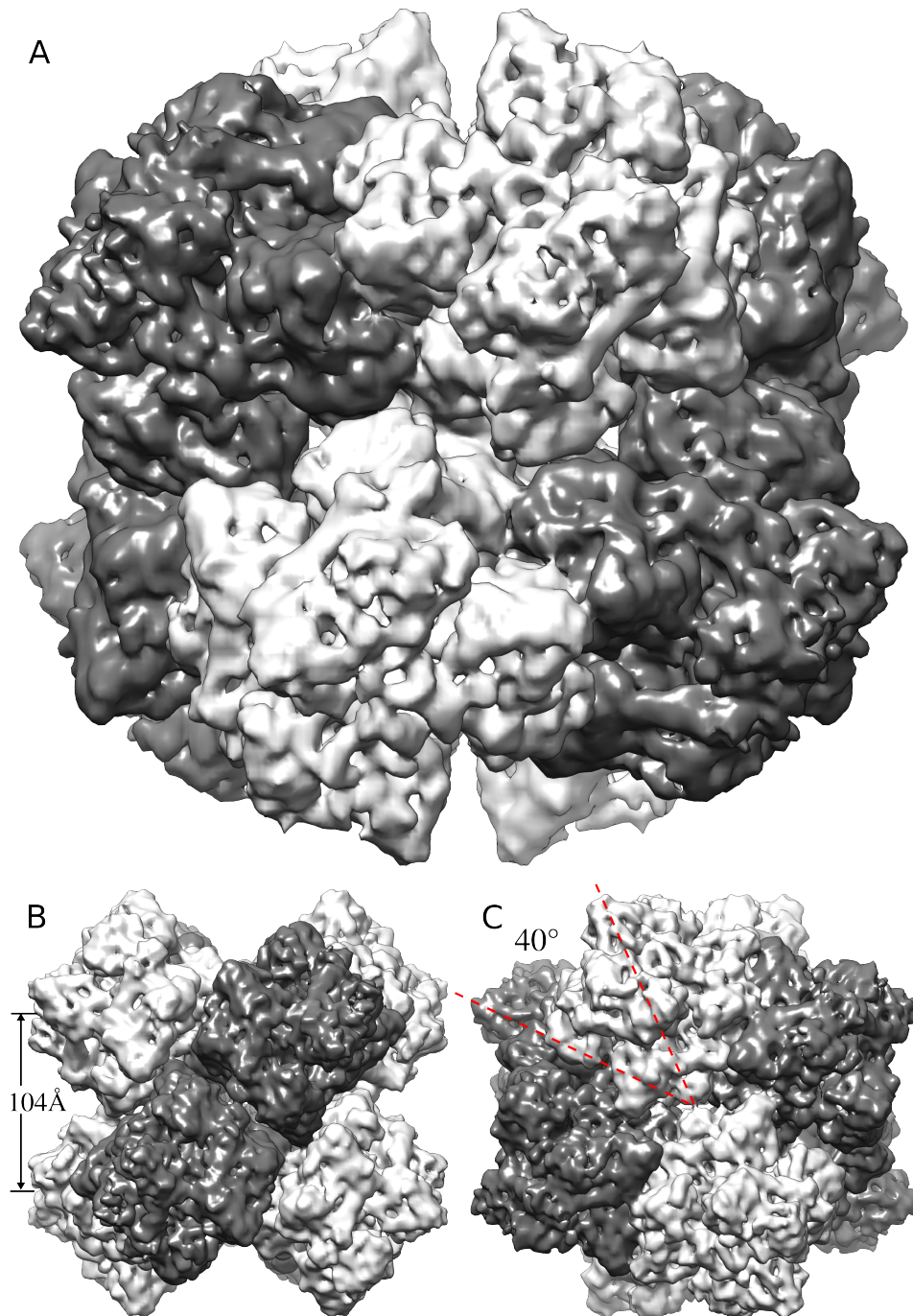
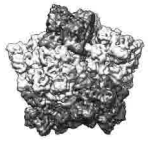


Abb. C.12: Finale 3D-Rekonstruktion des Hämocyanins von *L. polyphemus* (Auflösung: 7,1 Å). (A) Ansicht entlang der z -Achse. (B) Ansicht entlang der y -Achse, und der Abstand der 4x6mere im 8x6mer. (C) Ansicht entlang der x -Achse mit dem Staffelungswinkel von 40° (Verdrehung des vorderen 4x6mers zum interen 4x6mers).

Hexamer-Interface IIIA–IIIA, dass sich keine bzw. nur eine sehr schwache Brücke ausbildete (Abb. C.12 B und C.15). Aus diesem Grund wurden zwei weitere Rekonstruktionen nach dem gleichen Prinzip erzeugt, jedoch mit dem angepassten Parameter $-ir$ von 2 Pixeln und unter Ausschluss von Symmetrie (Kapitel C.3.4). Diese Rekonstruktionen zeigten eine deutlichere Brücke im Bereich des IIIA–IIIA-Kontakt, erreichten aber nur eine geringere Auflösung.

3.2 Beschreibung der 7,1 Å-Rekonstruktion

Das Hämocyanin von *L. polyphemus* besaß die Abmessungen von 260 Å entlang der x -Achse, 267 Å entlang der y -Achse und 254 Å entlang der z -Achse. Da sich das gesamte Molekül aus acht einzelnen Hexameren zusammensetzt, wurden diese untereinander genau vermessen. Es setzen sich jeweils zwei Hexamere zu einem 2x6mer zusammen und bilden somit die kleinste asymmetrische Untereinheit. Zwei dieser 2x6mere bilden ein 4x6mer und zwei dieser 4x6mere bilden zusammen das Hämocyanin von *L. polyphemus* (8x6mer). Die in Abb. C.10 gezeigten Klassensummen lassen diesen Aufbau schon erahnen. Für die weitere Vermessung wurde die Nomenklatur von Martin et al. (2007) übernommen.

Die Distanz zwischen zwei einzelnen 1x6meren innerhalb eines asymmetrischen 2x6mers beträgt 99 Å (Abb. C.13). Die beiden Hexamere sind um 105° gegen den Uhrzeigersinn verdreht. Dabei zeigt das eine Hexamer eine Drehung von 3° um die dreizählige Symmetrieachse des anderen Hexamers. Die dreizähligen Symmetrieachsen der beiden Hexamere bleiben fast parallel (3°). Die Distanz der beiden 2x6mere innerhalb des 4x6mer weicht um 9 Å gegenüber der Angabe von Martin et al. 2007 ab und entspricht 93 Å. Diese Entfernung wurde zwischen den Masseschwerpunkten (Zentroiden) der jeweiligen Hexamere gemessen. Die 2x6mere sind leicht verschoben (Abb. C.13 C), was in einer Dihexamerverschiebung (*flip-flop shift*) von 12 Å und einem Kippwinkel (*rocking angle*) von 14° resultierte (Abb. C.13 B). Die Distanz der 4x6mere im 8x6mer beträgt 104 Å und hat einen Staffelungswinkel (*staggering angle*) von 40° (siehe Abb. C.12 B und C). Die Zusammenfassung der Messwerte, zusammen mit den Ergebnissen aus Martin et al. (2007), sind in Tabelle C.4 aufgeführt.

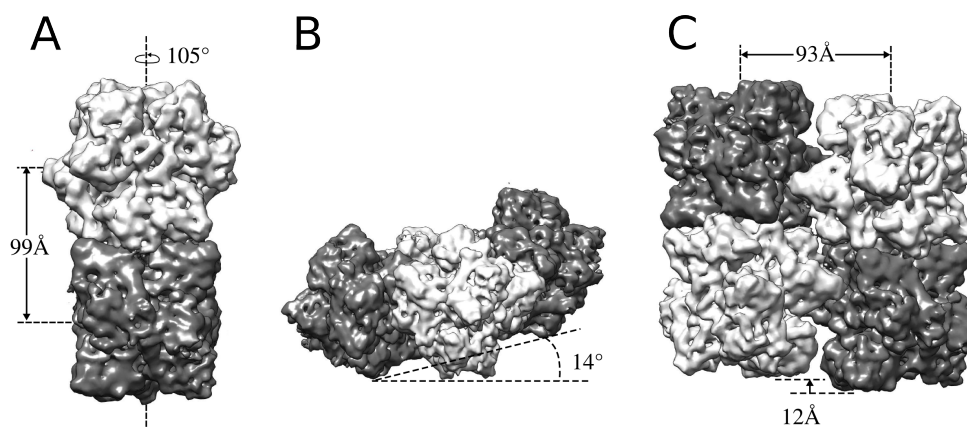


Abb. C.13: Vermessung der 3D-Rekonstruktion des *L. polyphemus*-Hämocyanins. (A) Darstellung der asymmetrischen Untereinheit (2x6mer). Die Zentren der Hexamere sind 99 Å entfernt, das obere Hexamer ist um 105° gegen den Uhrzeigersinn verdreht. (B) Kippwinkel (*rocking angle*) von 14° der 2x6mere innerhalb des 4x6mers. (C) Dihexamerverschiebung (*flip-flop shift*) zweier 2x6mere von 12 Å und 93 Å Abstand, gemessen zwischen den jeweiligen Masseschwerpunkten (Zentroiden).

Tab. C.4: Strukturelle Parameter des *L. polyphemus*-Hämocyanins. Zum Vergleich sind die Werte aus Martin et al. (2007) mit aufgeführt.

1x6mer im 2x6mer		2x6mers im 4x6mer			4x6mers im 8x6mer		Referenz
Distanz (Å)	Rotation (°)	Distanz (Å)	Dihexamerverschiebung (Å)	Kippwinkel (°)	Distanz (Å)	Staffelungswinkel (°)	
100 ± 3	105 ± 5	102 ± 3	14 ± 1	16 ± 2.5	106 ± 3	42 ± 2	Martin et al. (2007)
99 ± 2	105 ± 2	93 ± 2	12 ± 1	14 ± 3	104 ± 2	40 ± 2	diese Arbeit

3.3 Flexibles Fitting der Untereinheiten mit Direx

Um die Bindestelle zwischen den Hexameren auf atomarer Ebene untersuchen zu können, musste zuerst ein pseudoatomares Modell der Quartärstruktur erzeugt werden. Hier wurde als Startstruktur das pseudoatomare Modell von Martin et al. (2007) benutzt und die Hämocyanin-Untereinheiten einzeln mit dem Modul `fit in map` (Chimera) in die 7,1 Å-Rekonstruktion eingepasst. Im Anschluss wurde mit der Software `Direx` das flexible Fitting („flexibles Einpassen“) durchgeführt. Es wurde neben den Grundeinstellungen mit den in Tabelle C.5 gezeigten Werten gearbeitet.

Durch diese Einstellungen konnte die Korrelation im Messbereich von 7,7 bis

Tab. C.5: *Direx*-Parameter für das flexible *Fitting*

Parameter	Standard Wert	angewendeter Wert
<code>nsteps</code>	100	100
<code>annealing</code>	0	100
<code>den_strength</code>	0.4	0.2
<code>den_resid_range</code>	-1	20
<code>cur_map_dmin</code>	5.0	6.7
<code>cur_map_dmax</code>	200.0	400.0
<code>map_cv_dmax</code>	10.0	7.7
<code>expd_strength</code>	0.1	0.4

400 Å (*Cwork*-Wert) von 0.872 auf 0.937 erhöht werden, und im Bereich von 6,7 bis 7,7 Å (*Cfree*-Wert) von 0.198 auf 0.206. Der Anstieg beider Werte deutete darauf hin, dass damit noch kein Überfitten stattfand. Abb. C.14 zeigt ein einzelnes Hexamer der 7,1 Å-Rekonstruktion vor und nach dem flexiblen *Fitting*.

Zusätzlich zu den angepassten Parametern wurde eine Datei mit den Wasserstoffbrücken des atomaren Modells erzeugt und in den Prozess des flexiblen Fittings einbezogen. Dadurch wurde das Molekül zusätzlich stabilisiert, da es sonst in diesem Auflösungsbereich des *Cfree*-Wertes schnell zu einem Überfitten kommt. Zur Herstellung der Datei wurde `FindHBond` (*Chimera*) benutzt und mit dem Programm `hbond2direx` (siehe Kapitel 4.6.2) in ein für *Direx* lesbares Format umgeschrieben. Die Daten wurden von der Software als Einschränkungen (*restrains*) behandelt und gegenüber anderen Regeln bevorzugt.

3.4 Korrektur der 3D-Rekonstruktion im Bereich der IIIA–IIIA-Bindestelle

Wie in Abschnitt C.3.1 erwähnt, hatte die 7,1 Å-Rekonstruktion zu wenig Masse im Bereich der IIIA–IIIA-Bindestelle. Martin et al. (2007) identifizierten an dieser Bindestelle mehrere Aminosäuren, die sich für eine chemische Bindung eignen. Da die IIIA–IIIA-Bindestelle genau auf der Hauptsymmetrieachse der 3D-Rekonstruktion liegt, waren an dieser Stelle Artefakte möglich. Um dies genauer zu untersuchen, wurden drei 3D-Rekonstruktionen mit unterschiedlichen Einstellungen berechnet und die Ergebnisse an der IIIA–IIIA-Bindestelle verglichen.

In der 7,1 Å-Rekonstruktion wurde durch den Parameter `-ir=5` während der 2D-Ausrichtung der Einzelpartikel der innere Radius auf 5 Pixel definiert. Das

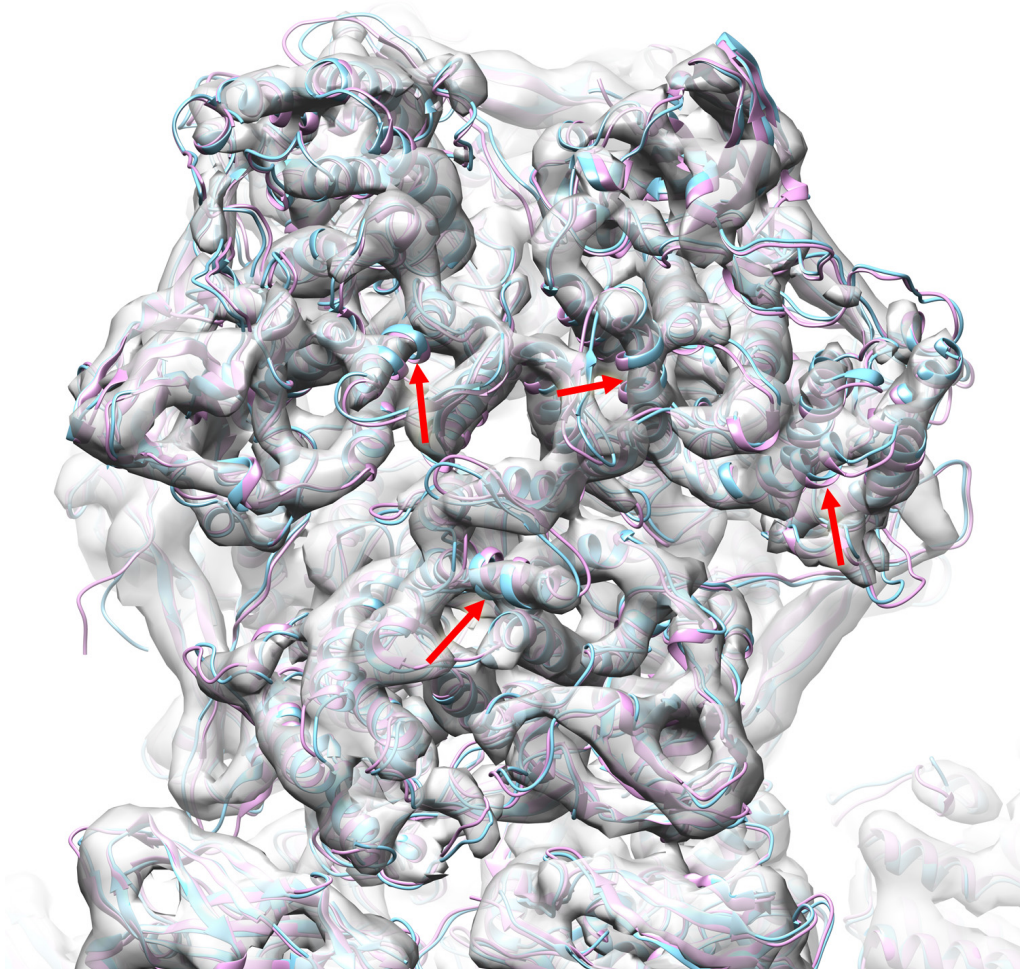
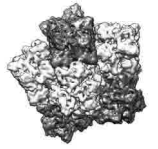


Abb. C.14: Effekt des flexiblen *Fittings* (Einpassen) mit *Direx*, gezeigt am Beispiel eines Hexamers der 7,1 Å-Rekonstruktion. In Rosa ist das Ausgangsmodell zu sehen (Korrelation mit 87,2%), und in Türkis das finale Modell (Korrelation von 93,7%). Die Pfeile deuten auf Helix-Bereiche, in denen man die Veränderung besonders gut sehen kann, indem die Sekundärstrukturen besser in der Masse sitzen.

bedeutet, dass die ersten 5 Pixel von der Symmetrieachse nicht in die Ausrichtung einbezogen werden. Ein Wert dieser Größe soll Artefakte verhindern (Penczek et al., 1992). Im Fall der IIIA–IIIA-Bindestelle wurde hier jedoch eine zu geringe Masse berechnet und es zeigt sich keine Verbindung zwischen den 4x6meren bei 70 % der angezeigten Masse (Abb. C.15).

Um mehr Masse und somit Artefakte im Bereich der Symmetrieachse zu provozieren, wurde der Datensatz erneut nach dem oben genannten Verfahren berechnet, jedoch mit einem inneren Radius von nur 2 Pixeln. Das Ergebnis dieser Rekonstruktion ist weitestgehend mit der 7,1 Å-Rekonstruktion vergleichbar, erreichte aber nur 7,4 Å Auflösung.

Eine genauere Betrachtung der IIIA–IIIA-Bindestelle ergab bei gleichem Schwellenwert deutlich mehr Masse (Abb. C.16) im Vergleich zur 7,1 Å-Rekonstruktion.

Um die Artefakte in der Nähe von Symmetrieachsen komplett auszuschließen wurde eine weitere Rekonstruktion ohne Symmetrie gerechnet (C1-Symmetrie). Hierzu wurde aufgrund der großen Anzahl an Referenzen im Fall der C1-Symmetrie auf den kompletten Datensatz (38 032 Partikel) zurückgegriffen und nicht nach dem Goldstandard-Verfahren gerechnet. Die finale Auflösung lag bei 10 Å nach 0.5-Kriterium. Als Referenz diente die 7,1 Å-Rekonstruktion, die auf 50 Å gefiltert wurde. Zusätzlich wurde sie im 3D-Raum gedreht und neu ausgerichtet, um eine komplette Neuorientierung der Einzelpartikel zu erzwingen.

Das Ergebnis (Abb. C.17) ist nicht direkt mit den anderen Rekonstruktionen vergleichbar, da es sich um eine geringere Auflösung handelt. Jedoch ist im Bereich der IIIA–IIIA-Bindestelle eine deutlich höhere Masse zu erkennen als bei den anderen Rekonstruktionen.

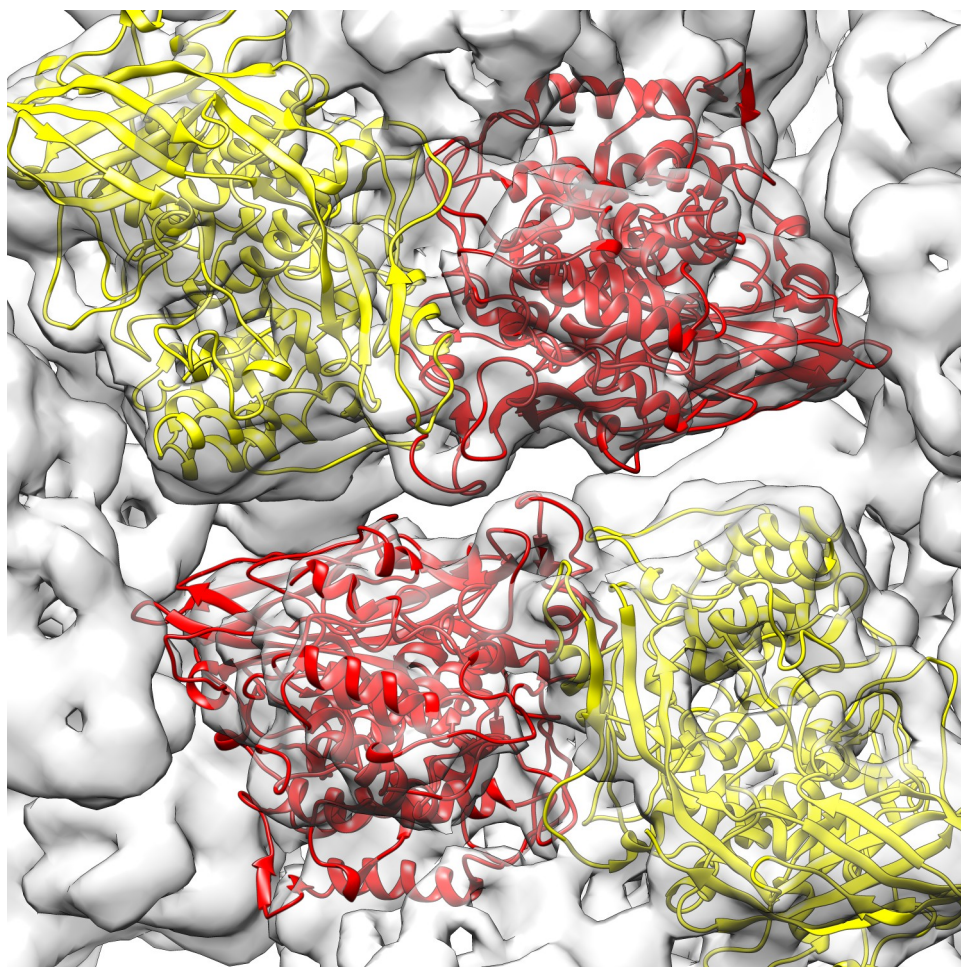
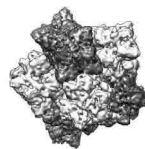


Abb. C.15: 7,1 Å-Rekonstruktion mit D2-Symmetrie von *L. polyphemus* mit Fokus auf die IIIA–IIIA-Bindestelle. Gerechnet wurde hier mit einem inneren Radius von 5 Pixel. Zwischen den beiden 4x6meren ist nicht genug Masse in der Rekonstruktion vorhanden.

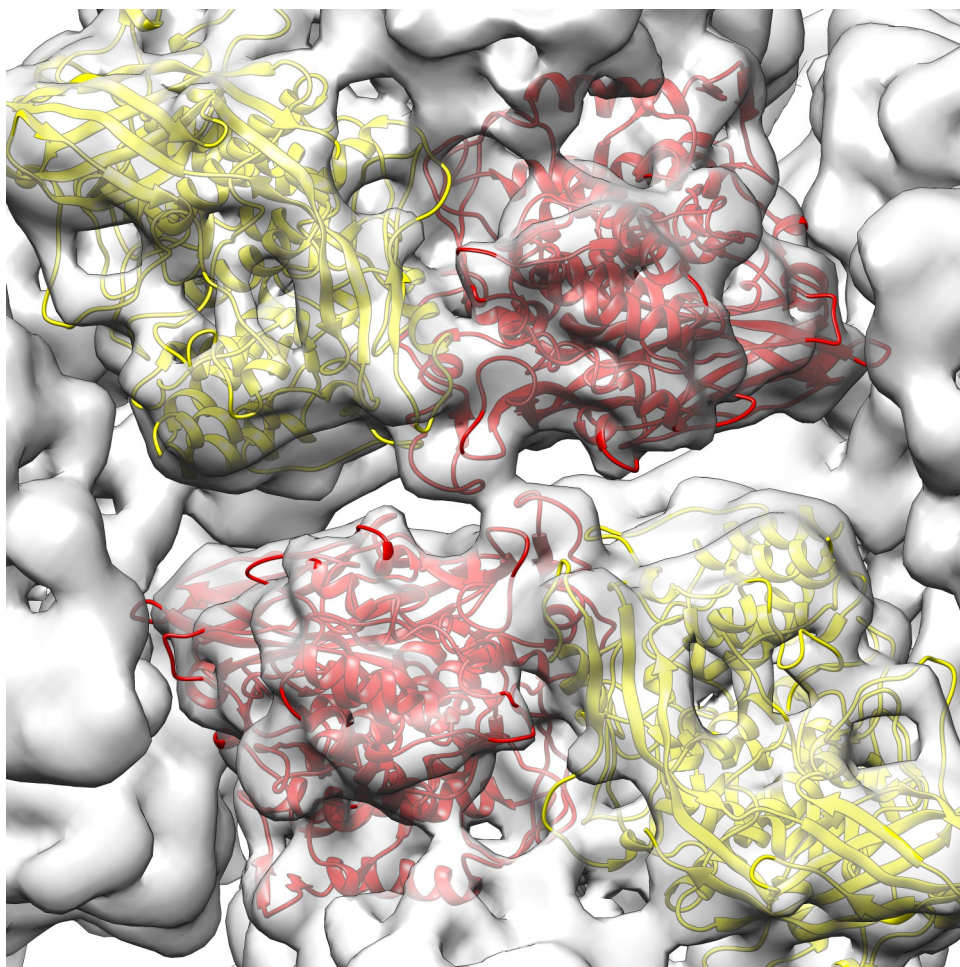


Abb. C.16: 7,4 Å-Rekonstruktion des *L. polyphemus*-Hämocyanins mit Fokus auf die IIIA–III A-Bindestelle. Durch den kleineren inneren Radius von 2 Pixeln bei der 2D-Ausrichtung konnte sich im Bereich der Symmetrieachsen mehr Masse akkumulieren. Somit blieb, bei gleichem Schwellenwert wie in Abb. C.15, im Bereich der Bindestelle eine Brücke erhalten.

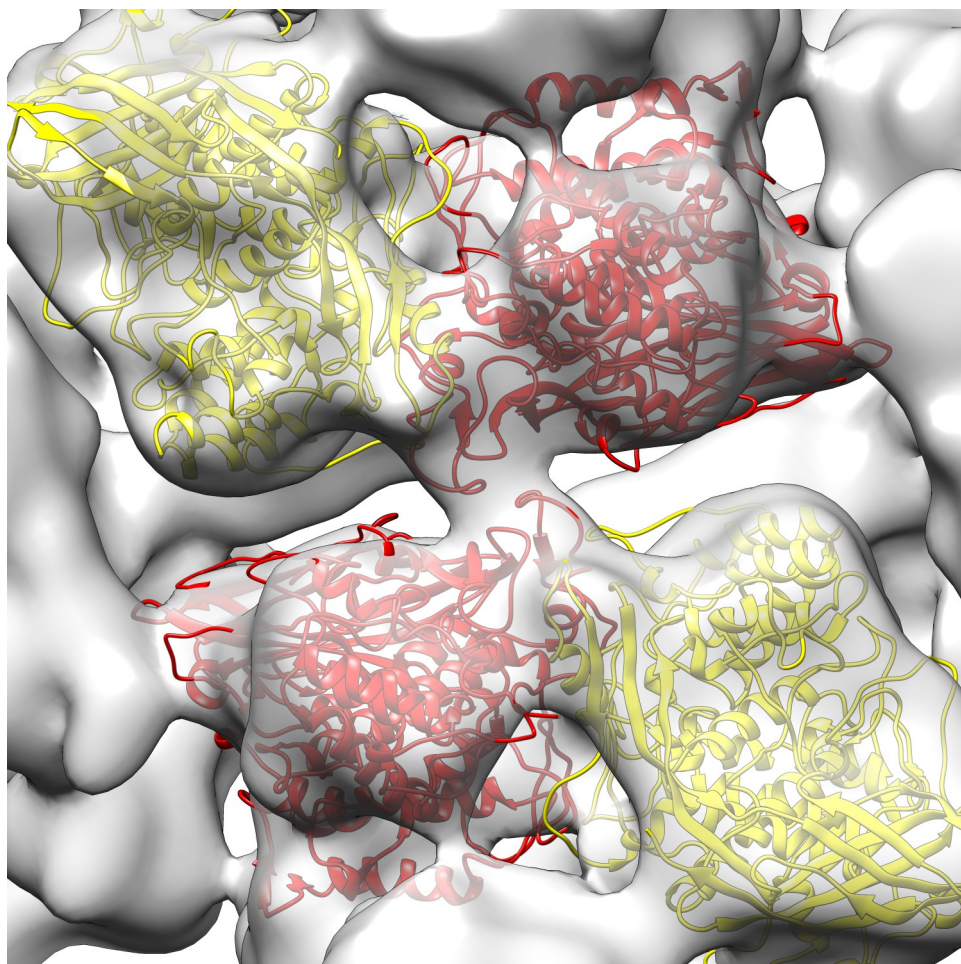
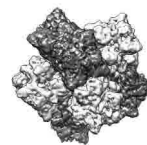


Abb. C.17: 10 Å-Rekonstruktion mit C1-Symmetrie von *L. polyphemus*-Hämocyanin mit Fokus auf die IIIA–IIIA-Bindestelle. Eine höhere Auflösung war mit dem vorhanden Datensatz beim Prozessieren ohne Symmetrie nicht zu erreichen. Die Brücke zwischen den beiden 4x6meren zeigt jedoch deutlich mehr Masse im Vergleich zu den anderen Rekonstruktionen.

3.5 Molekulardynamik-Simulation der IIIA–IIIA-Bindestelle

Die 3D-Rekonstruktionen ergaben im Bereich der Inter-Hexamer-Bindestelle IIIA–IIIA, die zwischen den beiden 4x6meren liegt, eindeutig eine Brücke. Für die Bindung geeignete Aminosäuren an dieser Stelle wurden von Martin et al. (2007) identifiziert: je zwei Kopien von D136, K137, H134 und H423 (Abb. C.18). Die 3D-Rekonstruktionen zeigten jedoch keine Details dieser hypothetischen Bindungen. Aus diesem Grund wurde von mir versucht, mithilfe von Molekulardynamik-Simulationen (MD-Simulationen) die Bindestelle und die beteiligten Aminosäuren genauer zu beschreiben.

Als Grundlage für die MD-Simulation diente das Ergebnis des flexiblen *Fittings* mit `Direx`. Hier wurden die beiden kontaktbildenden Untereinheiten vom Typ IIIA (hier IIIA₁ und IIIA₂ genannt) sowie die jeweils angrenzenden Untereinheiten vom Typ IIIB (hier IIIB₁ und IIIB₂ genannt) ausgeschnitten und mit dem `Gromacs`-Module `pdb2gmx` in ein `Gromacs` verarbeitbares Dateiformat umgewandelt. Da die Position der vier Untereinheiten durch die 7,1 Å-Rekonstruktion vorgegeben war, wurden alle Aminosäuren außerhalb eines Radius von 35 Å um die Kontaktstelle räumlich fixiert (Abb. C.19). Hierzu diente das von mir entwickelte Skript `restrain35angst.sh` (siehe Kapitel C.4.6). Es schrieb die Atomnummern aller Atome außerhalb des definierten Radius in eine Datei. Diese wurde vor dem Start der MD-Simulation an `Gromacs` übergeben, um die Atome in ihrer Bewegung einzuschränken (*restrain*). Die restlichen Atome konnten sich entsprechend der Simulation frei bewegen.

Zur Simulation wurden die vier Hämocyanin-Untereinheiten mittig mit 12 Å Abstand in eine quadratische Box positioniert. Als Kraftfeld für die Simulationen kam das `AMBER99SB-ILDN`¹ zum Einsatz, das sich für Proteinwechselwirkungen gut eignet (Lindorff-Larsen et al., 2012; Beauchamp et al., 2012). Wassermoleküle wurden mit dem einfachen `TIP3P`-Modell simuliert, das drei Ladungen vorsieht (Jorgensen et al., 1983, 1998): zwei gleiche positive Ladungen für die Wasserstoffatome und eine negative Ladung für das Sauerstoffatom, das gerade die positiven kompensiert. Die Box wurde danach mit Wassermolekülen befüllt (bereits minimiertes Standardmuster `spc216`). Im Anschluss wurden vereinzelte Wassermoleküle durch Ionen ausgetauscht und equilibriert. Hier kamen zwei unterschied-

¹`AMBER99SB-ILDN` ist in `Gromacs` implementiert und bildet die Grundeinstellung zum Berechnen der Kräfte und Energiepotentiale während der MD-Simulation.

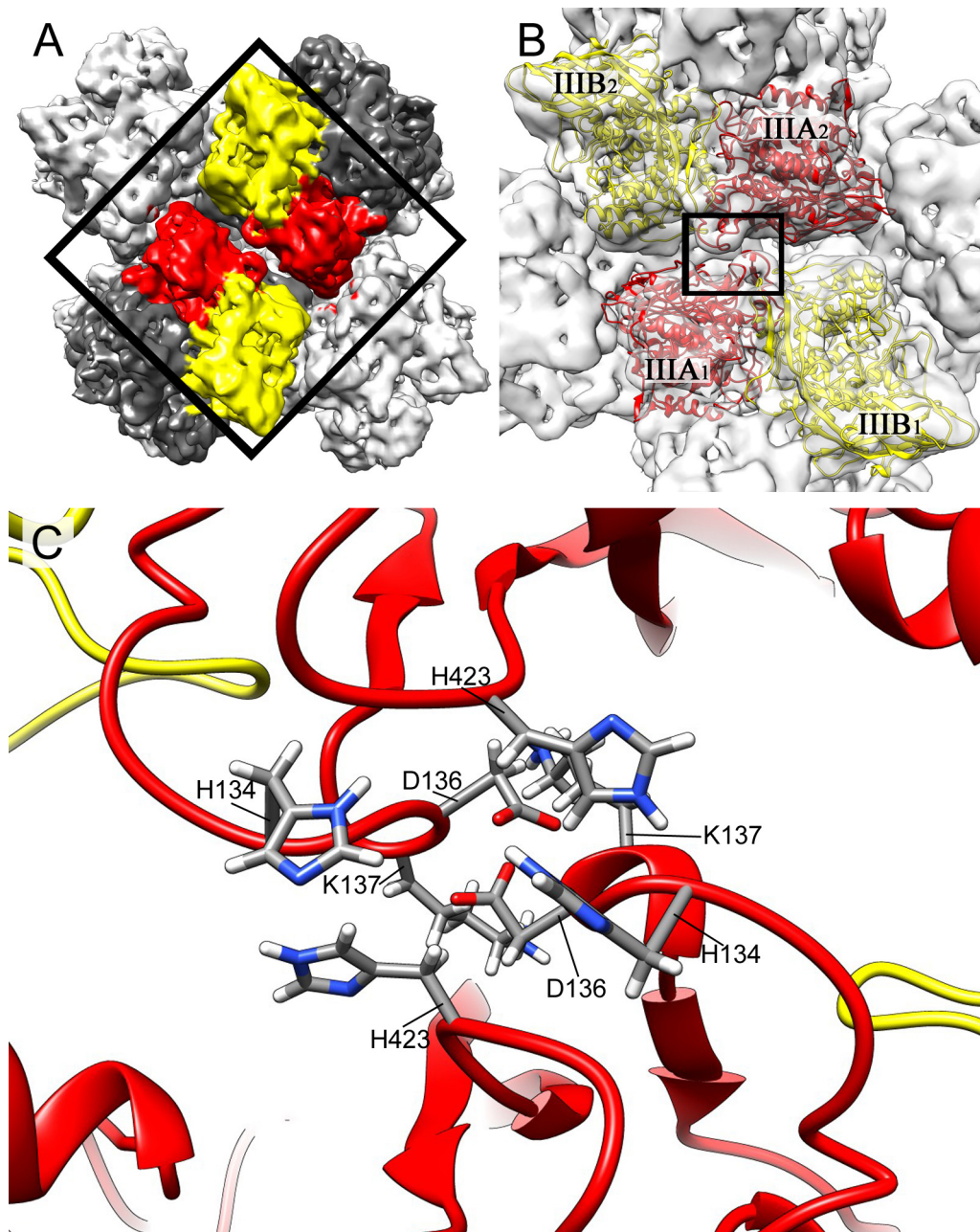
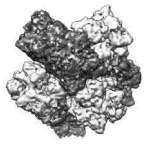


Abb. C.18: Position der IIIA-III A-Bindestelle im *L. polyphemus*-Hämocyanin mit der durch Direx eingepassten atomaren Struktur. (A) Blick entlang der y -Achse (siehe Abb. C.12). (B) Fokus auf schwarzen Rahmen aus (A) mit Beschriftung der Untereinheiten nach Martin et al. (2007). (C) Fokus auf schwarzen Rahmen aus (B) und Anzeige der nach Martin et al. (2007) identifizierten Aminosäuren H134, D136, K137 und H423.

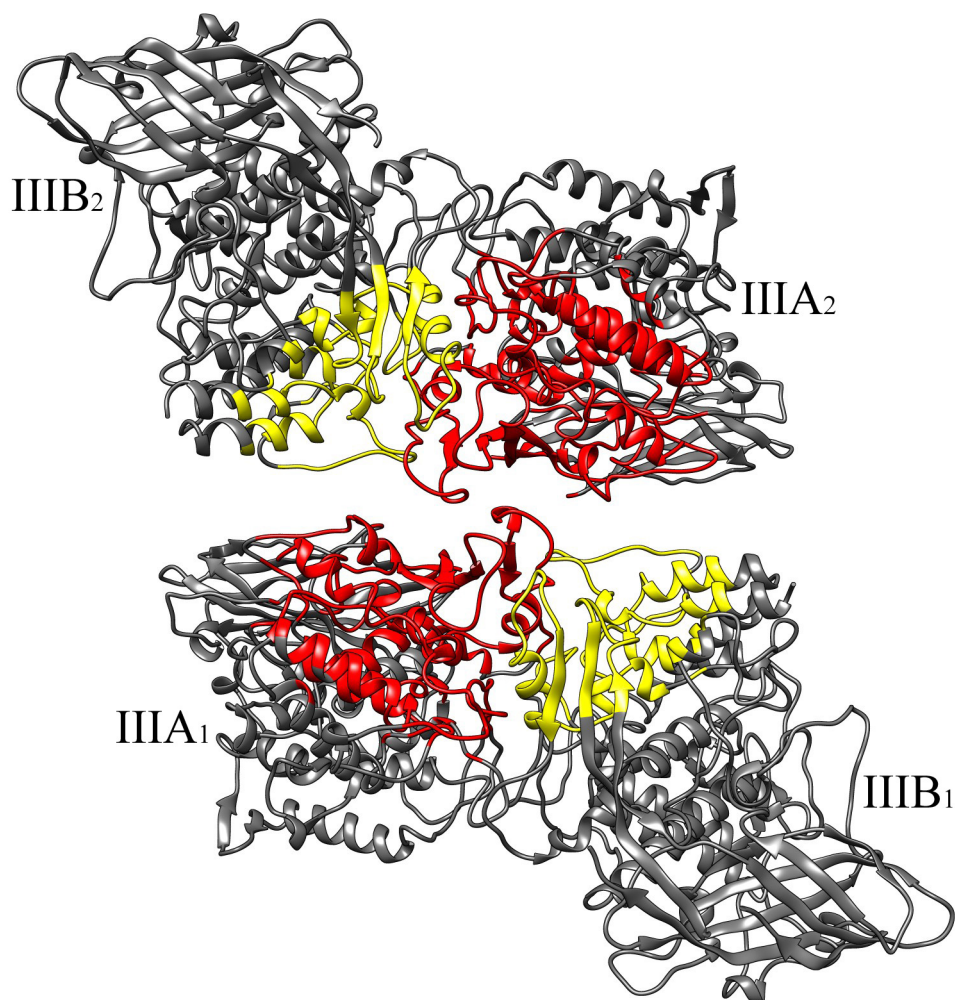
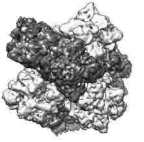


Abb. C.19: Die simulierten Hämocyanin-Untereinheiten der IIIA–IIIA-Bindestelle. Die dunkel eingefärbten Bereiche wurden dabei durch Einschränkungen (*restrains*) mit einer bestimmten Kraft im Raum fixiert. Die rot bzw. gelb eingefärbten Bereiche konnten sich unbehindert in der Simulation bewegen.



liche Zusammensetzungen des umgebenden Mediums zum Einsatz. Eine, indem 10 mM H₂O durch 10 mM NaCl ersetzt wurden (hier Puffer A genannt). Die andere, indem 10 mM H₂O durch 10 mM MgCl₂ ausgetauscht wurden (hier Puffer B genannt).

3.5.1 Zu analysierende Aminosäuren in der Bindestelle

Zur Analyse der Wechselwirkungen wurden jeweils die ersten 2 Nanosekunden (ns) einer Simulation entfernt, um Fehler durch thermisches Rauschen zu minimieren. Die Messung der Atomabstände wurden mit dem Gromacs-Modul `g_dist` durchgeführt. Man benötigte hierfür eine Datei (`index.ndx`) mit Gruppen von Atomen, zwischen denen die Abstände gemessen wurden.

Da die Entfernungsmessung zwischen zwei Atomen, aufgrund sich drehender Aminosäurenrest verfälscht werden kann, wurde unter Umständen eine Gruppe von mehreren Atomen an der Spitze ausgewählt (Abb. C.20 grün). In diesem Fall wurde nicht die Entfernung zwischen einzelnen Atomen, sondern zwischen den Massenschwerpunkten (Abb. C.20 magenta) bestimmt. Um die Gruppen zu definieren wurde das Programm `Chimera` eingesetzt und alle Atome der zu vermessenden Aminosäuren gleichzeitig ausgewählt. Danach wurde innerhalb von `Chimera` das Menü *Favorites* und dort die Python-Programmierungsumgebung `IDLE` ausgewählt. Dort wurde der Python-Programmcode aus Abb. C.34 (Kapitel 4.6.1) eingegeben und ausgeführt. Das Ergebnis ist eine Liste mit Gruppennamen und Atomnummern (Abb. C.21). Diese Daten wurden in die `index.ndx`-Datei aufgenommen und standen nun innerhalb des Moduls `g_dist` nur zur Auswahl für die Vermessung.

Zur Visualisierung der Daten wurden drei Skripte von mir entwickelt: `gromacsgraph.py`, `gromacshist.py` und `gromacsdisc.py` (Kapitel 4.6.1). Mit `gromacsgraph.py` werden die Abstände der Atome und Atomgruppen über die Zeit aufgetragen (Abb. C.22) und mit `gromacshist.py` in ein Balkendiagramm umgewandelt. Ein Balken repräsentiert den gerundeten Abstandswert in Ångström, die Höhe entspricht prozentual der Simulationszeit (also, wie lange die gemessenen Atome in diesem Abstand zueinander verweilten).

Um die Daten direkt miteinander vergleichen zu können, wurden die Balkendiagramme in ein kompaktes Scheibendiagramm umgewandelt. Hier entspricht die Fläche der Scheibe dem Prozentwert eines Balkens (exemplarisch in Abb.

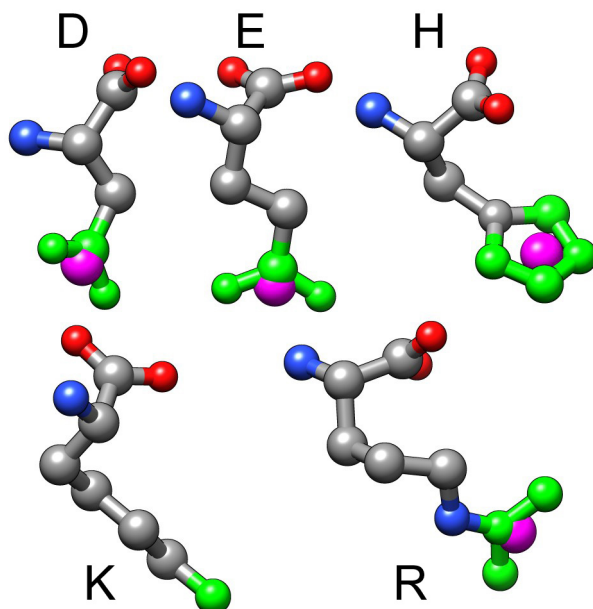


Abb. C.20: Ausgewählte Atome (grün) der Aminosäuren zum Messen der Abstände in der MD-Simulation. Bei mehreren Atomen wird der Masseschwerpunkt (magenta) benutzt.

```

1 [D514_C]
  28187 28188 28189
3 [H423_C]
  26723 26722 26717 26718 26719 26721 26720
5 [H134_A]
  2157 2158 2159 2153 2154 2155 2156
    
```

Abb. C.21: Von Chimera erstellte Atomgruppen zur Verwendung in Gromacs. Die Gruppennamen H134_A entsprach z. B. Histidin 134 auf Kette A (hier Untereinheit IIIA₁).

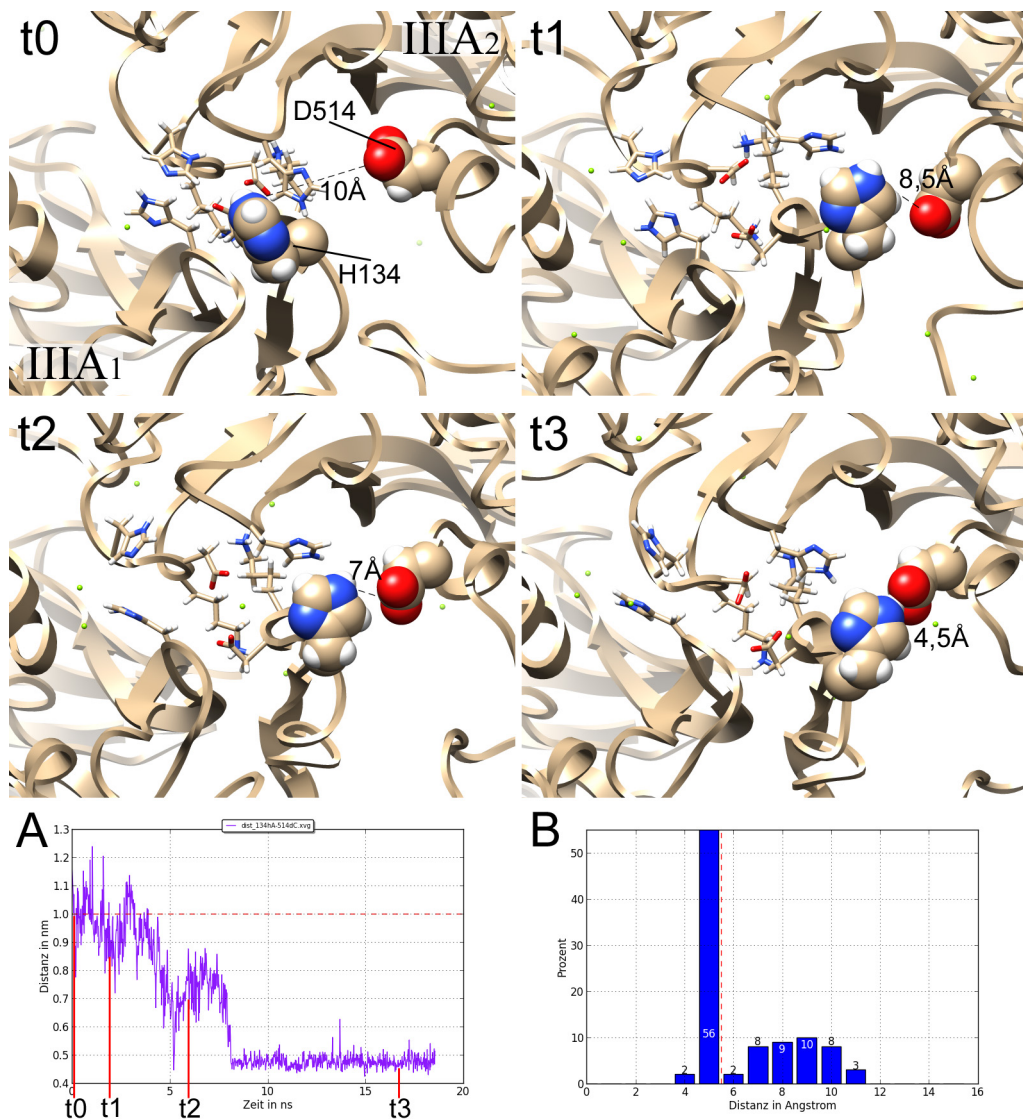
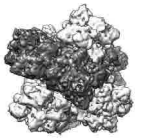


Abb. C.22: Analyse der MD-Simulation am Beispiel der Aminosäuren IIIA₁ H134 und IIIA₂ D514. (t₀) Zustand zum Zeitpunkt t₀ und Ergebnis des flexiblen *Fittings* mit *Direx*. Die Aminosäurenreste sind ca. 10 Å entfernt. (t₁) Zustand nach ca. 2 ns zum Zeitpunkt t₁. Die Aminosäurenreste haben sich bereits angenähert. (t₂) und (t₃) Zustand zum jeweiligen Zeitpunkt mit dem Abstand der Aminosäurenreste. (A) Verlauf der Distanz über die Zeit, die jeweiligen Zeitpunkte t₀ bis t₃ sind markiert. (B) trägt die Entfernungen in einem Balkendiagramm auf, gruppiert nach der Entfernung.

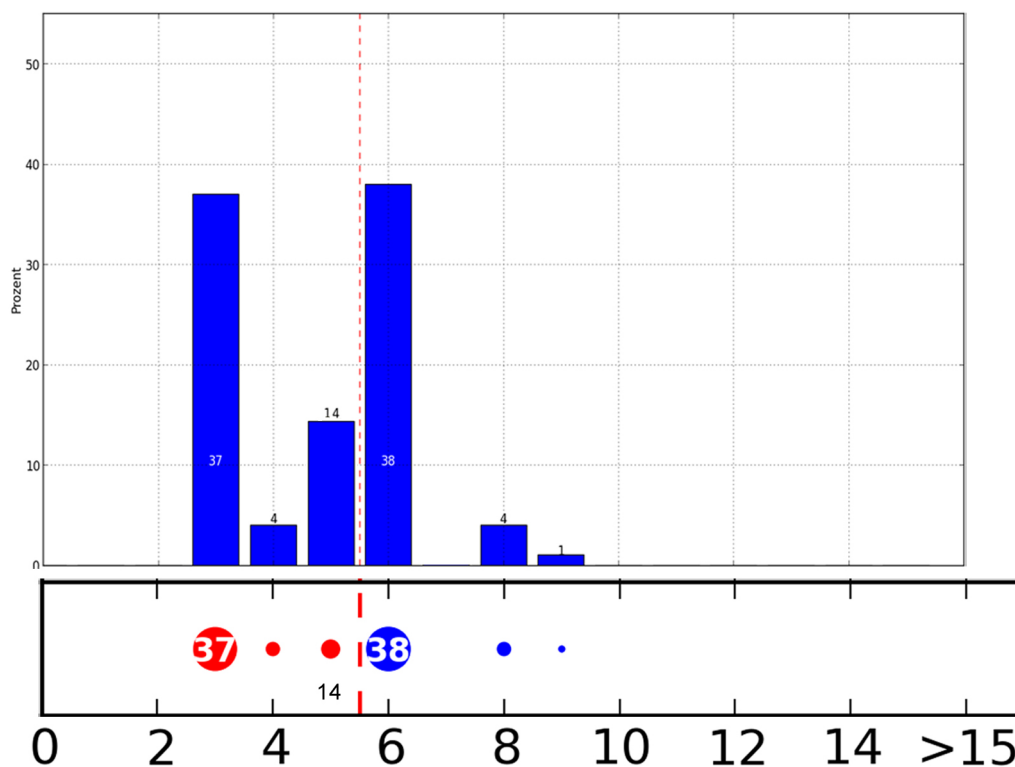


Abb. C.23: Zur Darstellung der in der MD-Simulation ermittelten Abstände werden die jeweiligen Säulen (oben) in Scheiben (unten) umgewandelt. Dabei entspricht die Fläche der Scheibe der Höhe des Balkens. Werte größer 30% sind in der Scheibe eingetragen, Werte zwischen 30% und 10% unter der Scheibe. Kleinere Werte wurden hier vernachlässigt. Die Scheiben unter 6 Å Abstände sind in rot dargestellt.

C.23 dargestellt). Scheiben im Bereich kleiner als 5,5 Å sind hier rot markiert. Zwecks Übersichtlichkeit wurde auf Werte unter 10% verzichtet. Größere Werte sind unter oder innerhalb der Punkte angegeben.

3.5.2 Puffer A mit H₂O und NaCl

Insgesamt wurden fünf MD-Simulationen über 20 Nanosekunden (ns) und eine MD-Simulation über 100 ns berechnet, und zwar mit gleicher Startkonfiguration aber unterschiedlichen Zufallszahlen (*random seed*). Die Daten wurden zu einem großen Datensatz zusammengefasst, wobei jeweils die ersten 2 ns entfernt wurden. Dadurch ergab sich eine Gesamtzeit von 188 ns. Die Untereinheiten befanden sich hier in einem Medium aus H₂O und 10 mM NaCl.



Beim Sichten der Ergebnisse konnten neben den von Martin et al. (2007) bereits vermuteten Aminosäuren (D136, K137, H134 und H423 auf der IIIA-Untereinheit) weitere identifiziert werden, die offenbar an der Kontaktbildung beteiligt sind, nämlich D138, E410, K426 und D514. Die Bindungsdauer sowie die Bindungspartner der Inter-Hexamer-Verbindungen finden sich in Abb. C.24 unter Puffer A. Die zugehörigen Graphen sind im Appendix ab Seite 136 aufgeführt.

Das Diagramm zeigt für Puffer A drei deutliche Inter-Hexamer-Verbindungen: IIIA₁ K426 ↔ IIIA₂ D138, IIIA₁ E410 ↔ IIIA₂ K137 und IIIA₁ K137 ↔ IIIA₂ E410. Eine vierte Verbindung IIIA₁ D136 ↔ IIIA₂ K137 besteht für ca. 20% der Simulationszeit. Aminosäure IIIA₂ K137 war somit über 50% der Zeit mit Aminosäuren der Untereinheit IIIA₁ verbunden.

Zusätzlich zu den Inter-Hexamer-Verbindungen (IIIA₁ ↔ IIIA₂) wurden ebenfalls einige Intra-Hexamer-Verbindungen untersucht und in Diagrammen zusammengefasst (Abb. C.25). Hier ist auffällig, dass sowohl in IIIA₁, als auch in IIIA₂, die Aminosäure K426 sehr lange mit E410 der gleichen Untereinheit verbunden war. Verglichen mit Abb. C.24 scheint die Aminosäure IIIA₁ K426 zwischen den Bindungspartnern IIIA₁ E410 und IIIA₂ D138 häufig zu wechseln.

In Abb. C.26 wird ein beispielhafter Zustand dargestellt. Abb. C.26 (A) zeigt die Inter-Hexamer-Verbindungen der in Abb. C.24 aufgeführten Aminosäuren. Abb. C.26 (B) zeigt neben den Intra-Hexamer-Verbindungen die entsprechenden Aminosäuren ohne passenden Partner innerhalb der 6 Å-Grenze.

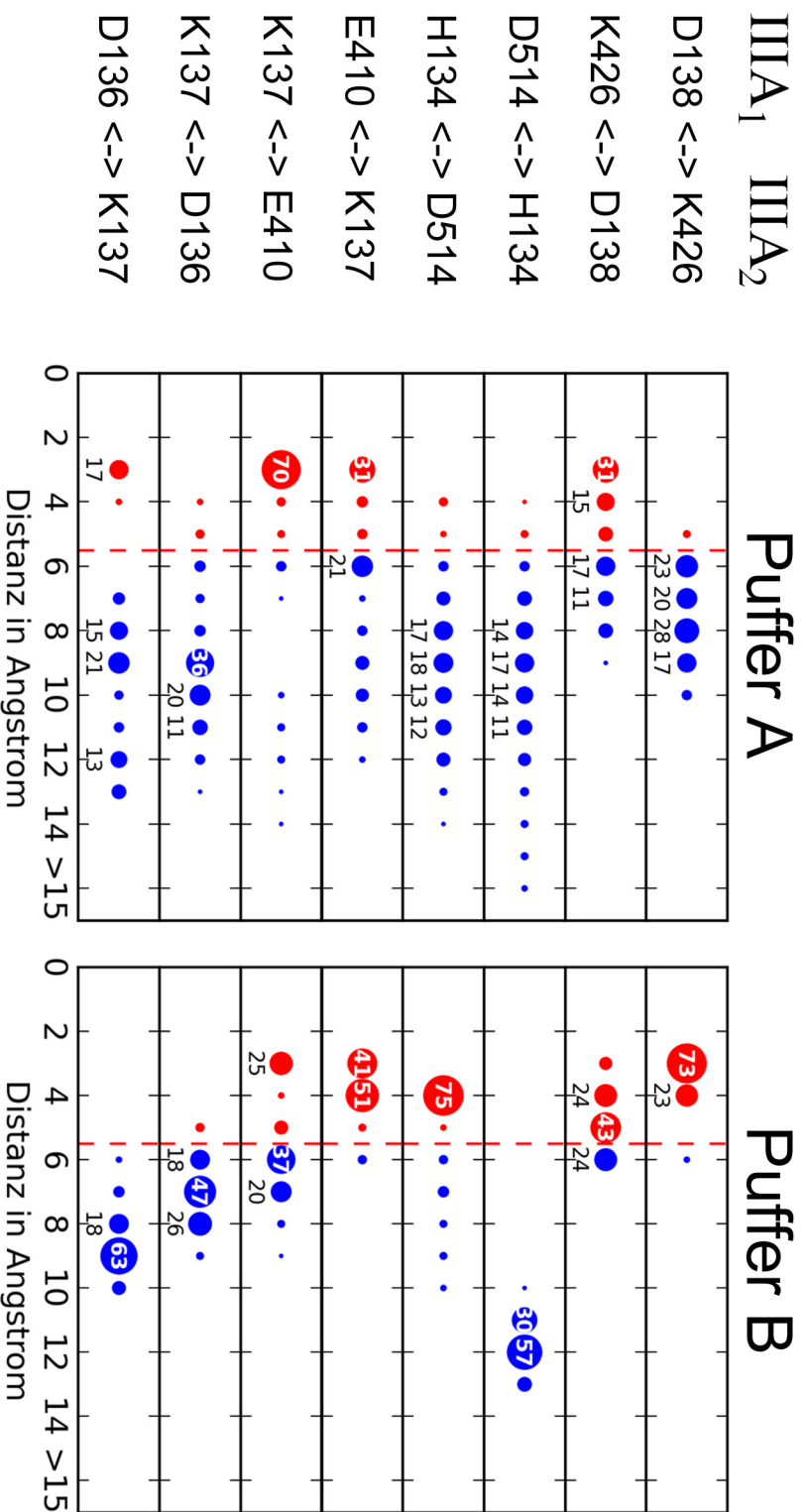


Abb. C.24: Ergebnisse der MD-Simulation der Bindestelle IIIA-III_A, dargestellt als Scheibendiagramm. Die Potentielle Bindungsdauer zwischen den Aminosäuren ist in Prozent dargestellt. Dabei bestanden bei Puffer A drei deutliche Verbindungen zwischen den Aminosäuren IIIA₁ K426 ↔ IIIA₂ D138, IIIA₁ E410 ↔ IIIA₂ K137 und IIIA₁ K137 ↔ IIIA₂ E410. Für Puffer B ergaben sich zusätzlich zwei weitere Verbindungen: IIIA₁ D138 ↔ IIIA₂ K426 und IIIA₁ H134 ↔ IIIA₂ D514.

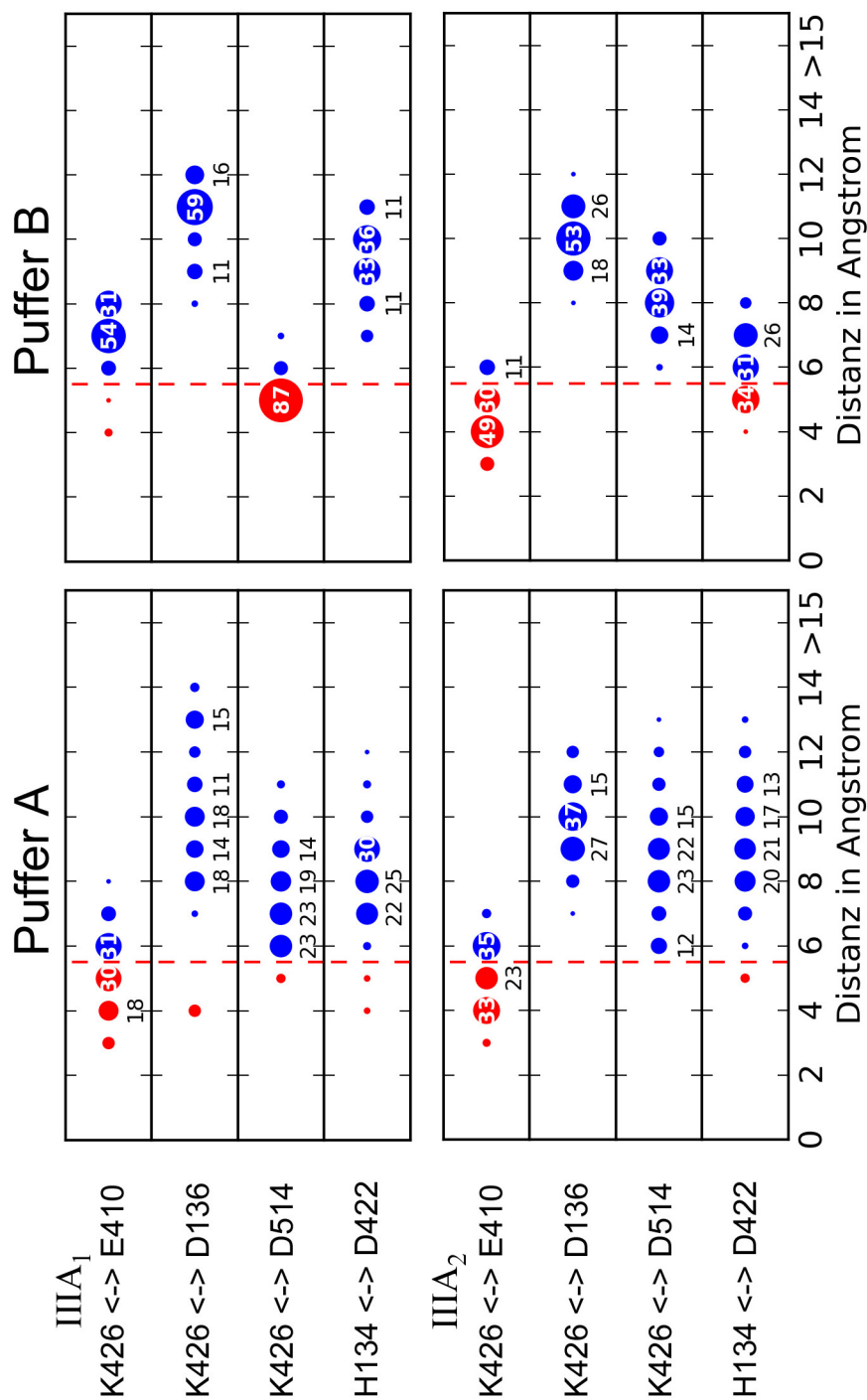
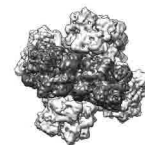


Abb. C.25: Ergebnisse der MD-Simulation der Intra-Hexamere-Bindestellen, dargestellt als Scheibendiagramm. Die Potentielle Bindungsdauer zwischen den Aminosäuren ist in Prozent dargestellt.

Innerhalb der IIIA₁-Untereinheit (oben) ergab sich eine Verbindung für K426 ↔ E410 in Puffer A und K426 ↔ D514 in Puffer B. Innerhalb der IIIA₂-Untereinheit (unten) ergaben sich in Puffer A und Puffer B zusätzlich die Verbindungen K426 ↔ E410 und in Puffer B eine weitere Verbindung K426 ↔ D514.

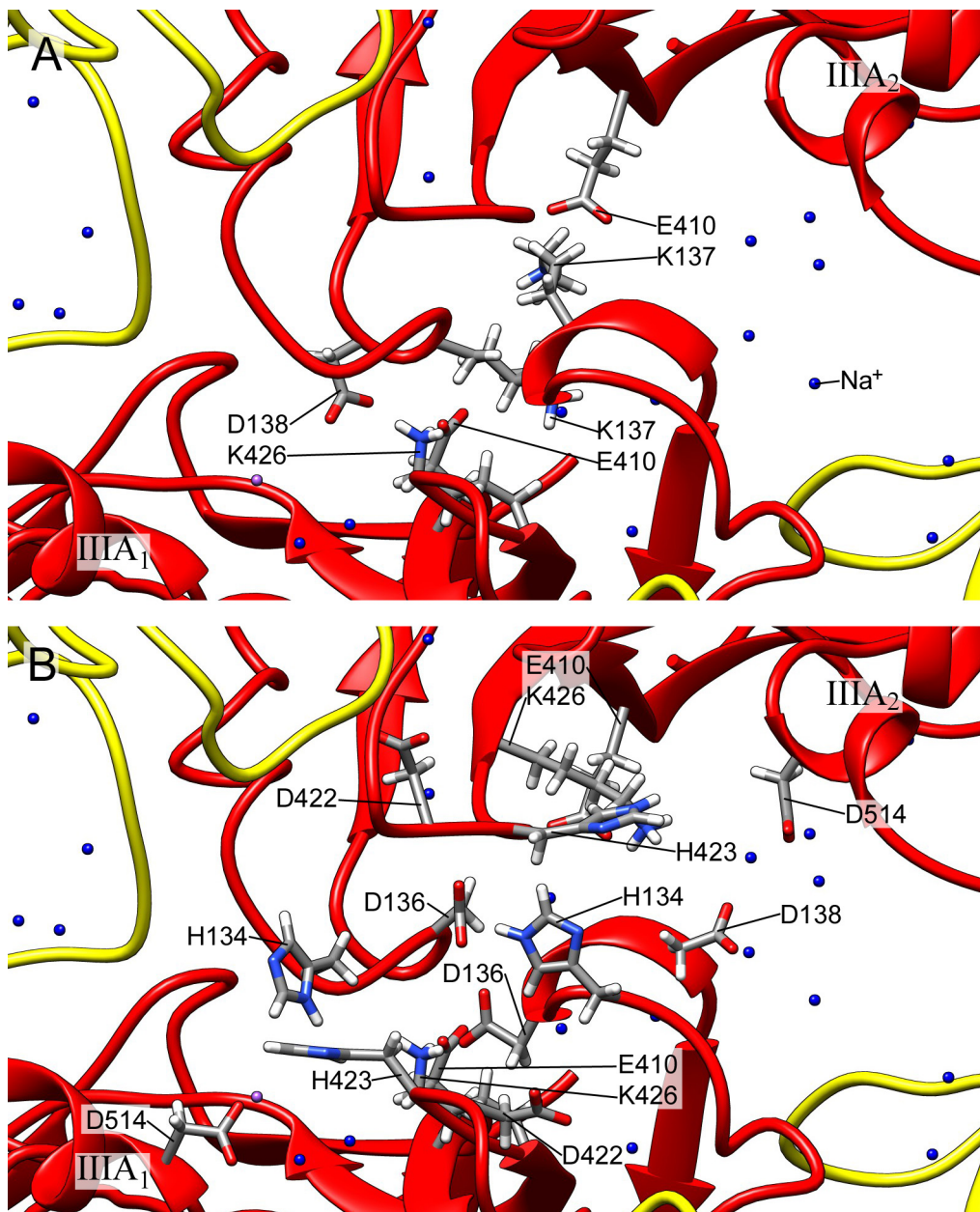


Abb. C.26: Beteiligte Aminosäuren der IIIA-III A Bindestelle in Puffer A. (A) Aminosäuren der Inter-Hexamer-Verbindungen paarweise markiert (K426 ↔ D138, E410 ↔ K137 und K137 ↔ E410). (B) Aminosäuren der Intra-Hexamer-Verbindungen paarweise markiert (E410 ↔ K426 und E410 ↔ K426), sowie die nicht bindenden Aminosäuren.



3.5.3 Puffer B mit H₂O, NaCl und MgCl₂

Um den Effekt von MgCl₂ zu testen, wurde zu Puffer A 10 mM MgCl₂ hinzugefügt (hier Puffer B genannt) und eine MD-Simulation angefertigt. Im Rahmen dieser Arbeit wurden ca. 26 ns berechnet, weitere Simulationen sind geplant. Nach dem Entfernen der ersten 2 ns verblieben 24 ns zur Analyse.

Wie die Scheibendiagramme zeigen (Abb. C.24 Puffer B), konnten die Aminosäuren D138, E410, K426 und D514 als an der Inter-Hexamer-Bindung beteiligt identifiziert werden. Zusätzlich zu den bereits in Puffer A beschriebenen Verbindungen $\text{III A}_1 \text{ K426} \leftrightarrow \text{III A}_2 \text{ D138}$, $\text{III A}_1 \text{ E410} \leftrightarrow \text{III A}_2 \text{ K137}$ und $\text{III A}_1 \text{ K137} \leftrightarrow \text{III A}_2 \text{ E410}$ treten in Puffer B zwei weitere Verbindungen auf, und zwar $\text{III A}_1 \text{ D138} \leftrightarrow \text{III A}_2 \text{ K426}$ und $\text{III A}_1 \text{ H134} \leftrightarrow \text{III A}_2 \text{ D514}$. Die zugehörigen Graphen sind im Appendix auf Seite 148 aufgeführt.

Bei den Intra-Hexamer-Kontakten ergaben sich wie in Puffer A die Verbindung $\text{III A}_2 \text{ K426} \leftrightarrow \text{III A}_2 \text{ E410}$, jedoch nicht für $\text{III A}_1 \text{ K426} \leftrightarrow \text{III A}_1 \text{ E410}$ (Abb. C.25). Dagegen existiert in Puffer B die Verbindung $\text{III A}_1 \text{ K426} \leftrightarrow \text{III A}_1 \text{ D514}$.

In der 3D-Ansicht wird ein beispielhafter Zustand in Abb. C.27 dargestellt. Abb. C.27 (A) zeigt die Inter-Hexamer-Verbindungen der in Abb. C.24 aufgeführten Aminosäuren. Abb. C.27 (B) zeigt neben den Intra-Hexamer-Verbindungen die entsprechenden Aminosäuren ohne passenden Partner innerhalb der 6 Å-Grenze.

3.6 Dichtekarte aus MD simulieren

Zum direkten Vergleich der 3D-Rekonstruktion wurde aus den 188 ns der Simulation alle 0,1 ns eine PDB-Datei extrahiert. Aus diesen PDB-Dateien wurde im Anschluss jeweils eine 3D-Rekonstruktion simuliert und alle Einzelvolumen zu einem einzigen Gesamtvolumen addiert. Hierfür diente das Script `md2volume.sh` (Kapitel 4.6.1), das diese Schritte automatisiert durchführt. Das Ergebnis wurde danach auf 7 Å (Abb. C.28) und 10 Å (Abb. C.29) gefiltert.

Dabei ähnelt der Bereich der IIIA–IIIA-Bindestelle am ehesten der 10 Å-Rekonstruktion (C1) (Abb. C.17). Dies ergibt Sinn, denn das in der MD-Simulation entstandene Inter-Hexamer-Interface III A_1 – III A_2 ist asymmetrisch.

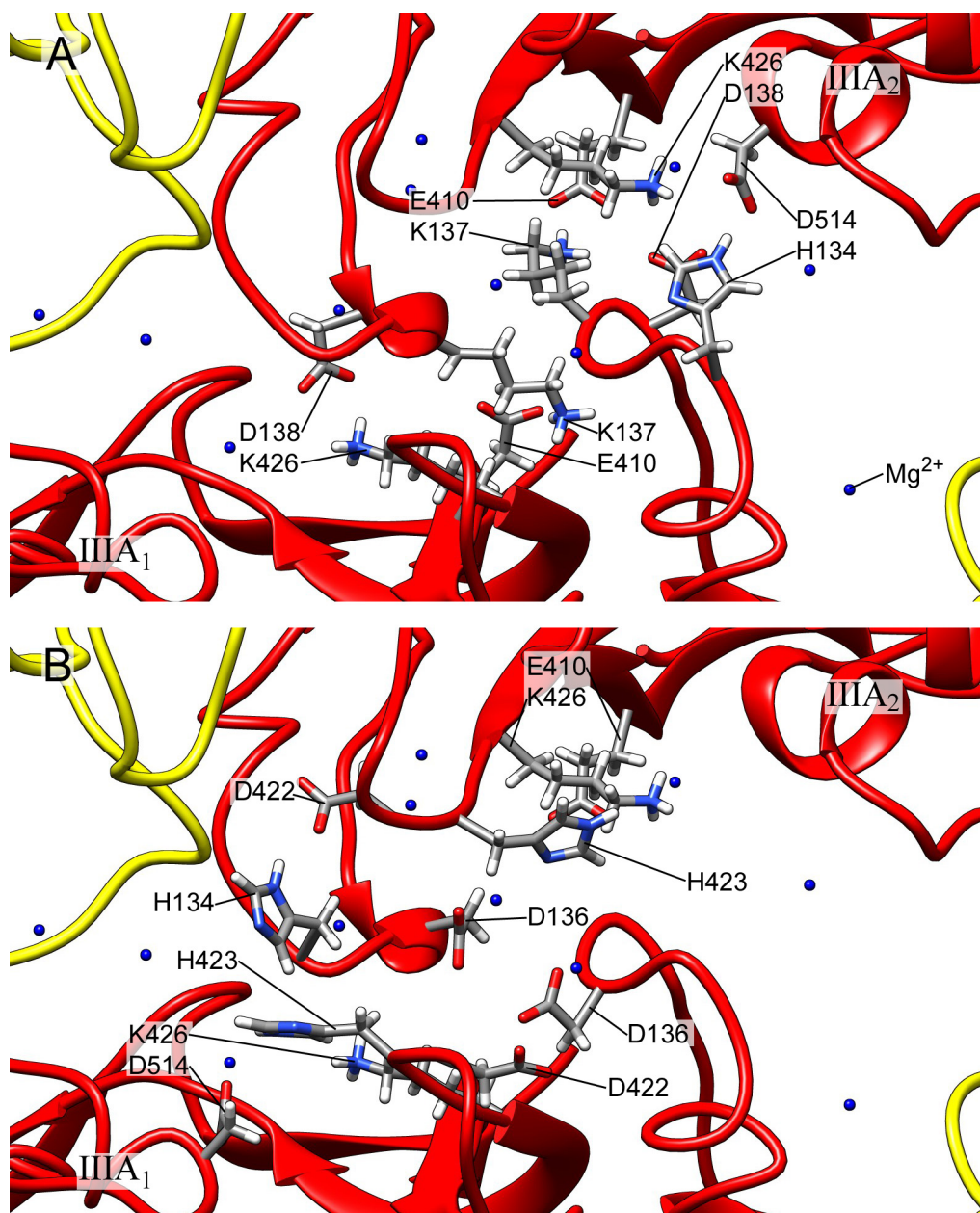


Abb. C.27: Beteiligte Aminosäuren der IIIA–III A Bindestelle in Puffer B. (A) Aminosäuren der Inter-Hexamer-Verbindungen paarweise markiert (K426 ↔ D138, E410 ↔ K137, K137 ↔ E410, D138 ↔ K426 und H134 ↔ D514). (B) Aminosäuren der Intra-Hexamer-Verbindungen paarweise (D514 ↔ K426 und K426 ↔ E410) sowie die nicht bindenden Aminosäuren einzeln markiert.

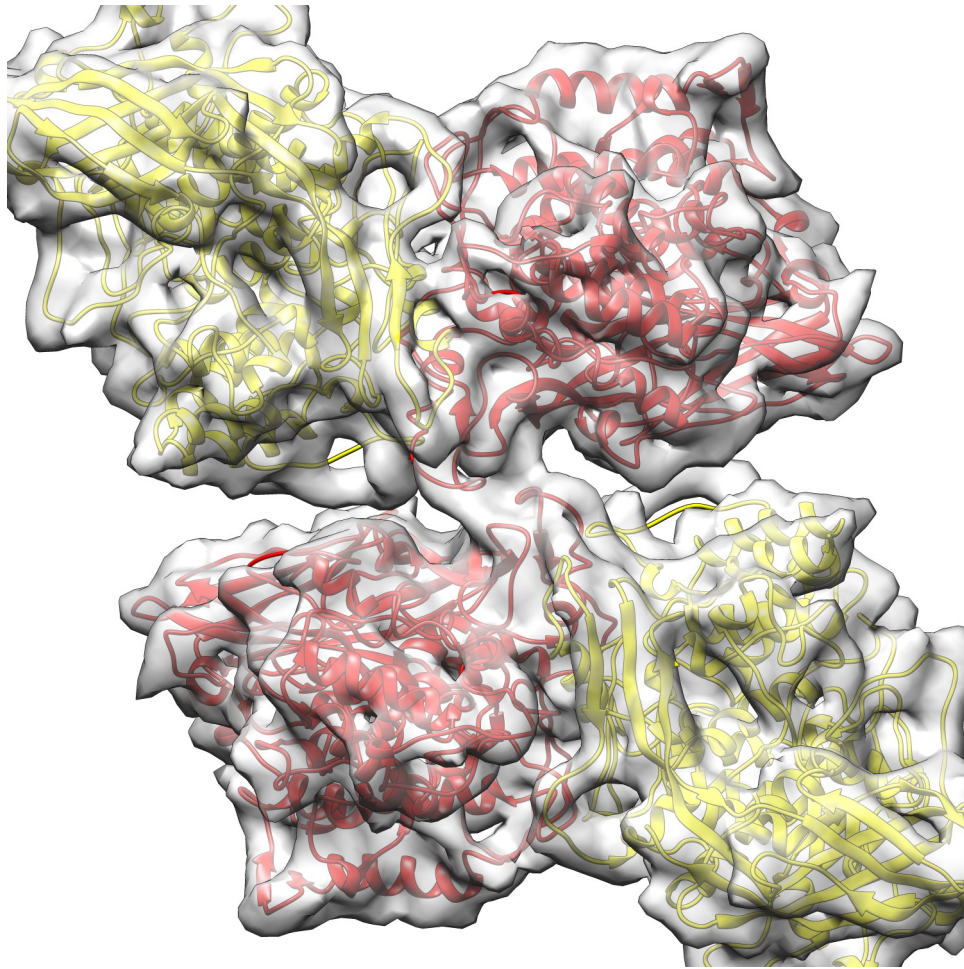
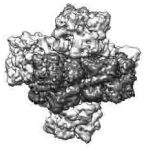


Abb. C.28: Dichtekarte der IIIA-III A-Bindestelle, berechnet aus den MD-Simulationen und auf 7 Å gefiltert.

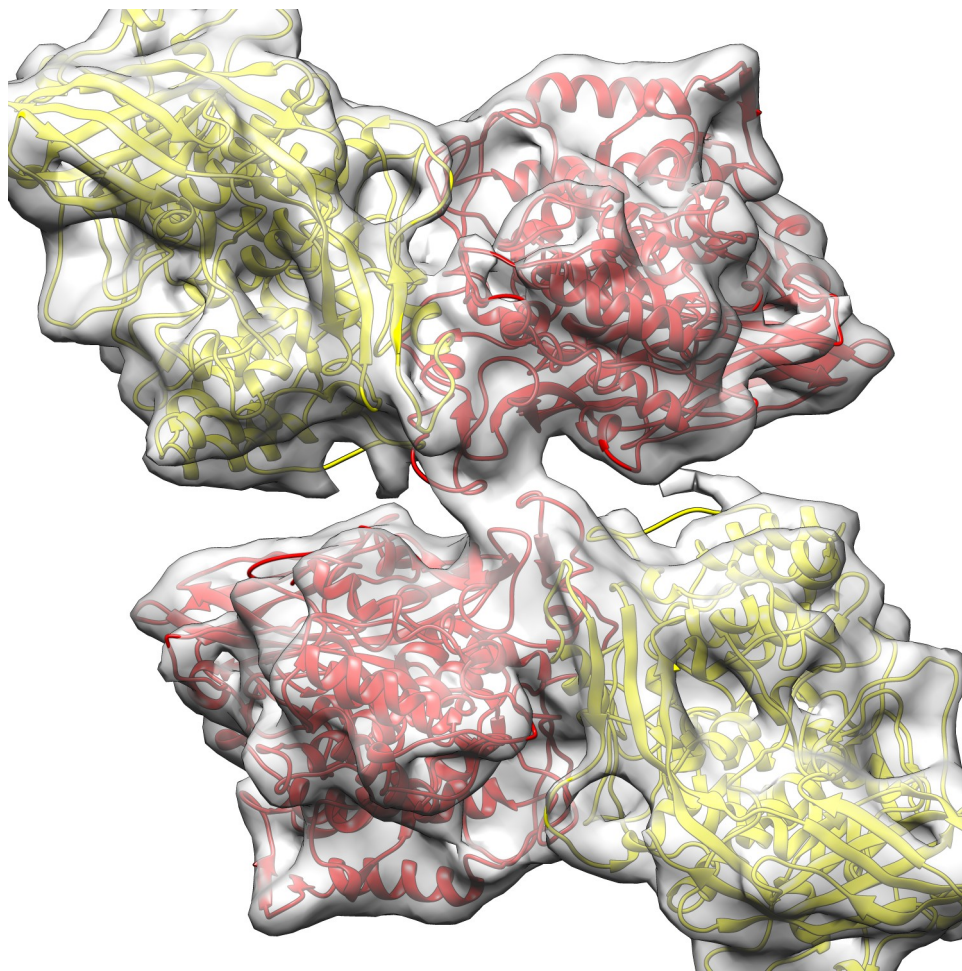


Abb. C.29: Dichtekarte der IIIA-III A-Bindestelle, berechnet aus den MD-Simulationen und auf 10 Å gefiltert.



4 Entwicklung einer grafischen Benutzeroberfläche und neuer Software-Tools

Im Rahmen meiner Arbeit entwickelte und implementierte ich neue Software-Tools, die die Prozesse während der 3D-Rekonstruktion erleichterten und beschleunigten. Um diese (hauptsächlich konsolenbasierten) Programme den Benutzern komfortabel zur Verfügung zu stellen, wurde eine grafische Benutzeroberfläche (*graphical user interface*) (GUI) entwickelt. Die GUI sollte übersichtlich und leicht erweiterbar sein. Die Tools wurden in benutzerdefinierte Kategorien eingeteilt und bildeten jeweils ein spezifisches Formular. Dieses wurde mit entsprechenden Daten gefüllt und in der Konsole gestartet. Die von mir entwickelten Software-Tools werden in Kapitel C.4.6 aufgeführt.

4.1 Die grafische Benutzeroberfläche MDEMTools

Um die GUI zu starten, meldet man sich mit einem `Terminal`-Programm auf das Rechen-Cluster der Arbeitsgruppe an und startet mit dem Befehl `MDEMTools` die GUI (Abb. C.30).

Die Menüstruktur entsteht beim Starten der GUI dynamisch; sie ist also abhängig von den eingelesenen Konfigurationsdateien (siehe Abschnitt C.4.2). Dabei ist das Menü *File* statisch und bietet nur die Option, die Oberfläche wieder zu schließen. Die Übrigen Menüpunkte entstehen entsprechend der Konfigurationsdateien. Unterhalb der Menüs befindet sich das Eingabeformular, mit einem statischen Feld zur Auswahl des Arbeitsverzeichnisses. Die restlichen Eingabefelder sind abhängig vom ausgewählten Werkzeug und passen sich entsprechend an.

Das in Abb. C.30 gezeigte Formular entspricht der Konfigurationsdatei in Abb. C.31. Zu erreichen ist es über die Kategorie *Appion* (Abb. C.32) und durch Auswahl des Werkzeugs `Appion2Relion`. Dadurch wird der entsprechende Titel und die Beschreibung sowie die benötigten Formularfelder angezeigt. Die Eingabefelder besitzen die Option auf einen Hinweistext (siehe Abb. C.30), der erst erscheint, wenn der Mauszeiger über dem Eingabefeld positioniert wird. Dieser Hinweistext sowie die Beschreibung, der Titel und die entsprechende Kategorie werden in der Konfigurationsdatei definiert.

Nach den Eingabefeldern folgt der Startknopf (*Execute*), um das eigentliche

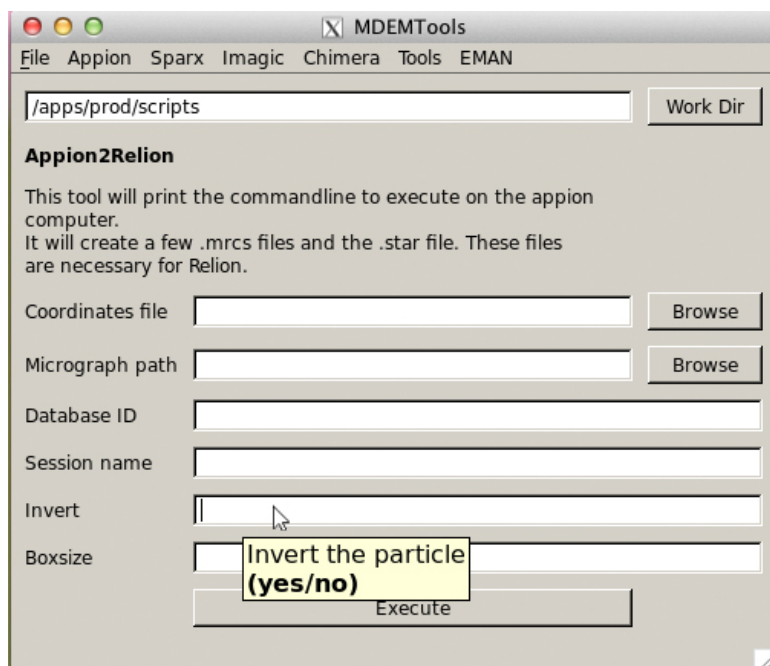


Abb. C.30: Grafische Benutzeroberfläche mit der in Abb. C.31 gezeigten Konfigurationsdatei. Ein Hinweistext erscheint zum jeweiligen Eingabefeld, wenn man mit dem Mauszeiger darüber fährt.

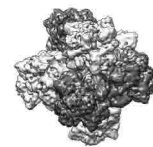
Programm auszuführen. Die Ausgabe geschieht im `Terminal`-Fenster. Dazu werden die eingegebenen Parameter an das entsprechende Werkzeug übergeben und gestartet.

4.2 Die Konfigurationsdatei

Bei den Konfigurationsdateien handelt es sich um einfache Textdateien, die mit einem Texteditor erstellt werden können. Um von der GUI erkannt zu werden, müssen sie auf dem Rechen-Cluster mit der Dateiendung `.cfg` im Ordner `/apps/prod/scripts/MDEMPugins` abgespeichert werden.

Der Aufbau der Konfigurationsdatei (siehe Abb. C.31) ist aufgeteilt in einen Kopf- und einen Formularbereich. Der Kopfbereich wird durch die Zeile `[HEAD]` eingeleitet und definiert die allgemeinen Programminformationen, wie z. B.:

- (*category*) passende Kategorien
- (*name*) Name des Programms



```
[HEAD]
2 category Appion
  name Appion2Relion
4 tip Convert your Appion project into a Relion project with .star file etc.
  command /apps/prod/scripts/appion2relion.py %s %s %s %s %s %s
6
  description This tool will print the commandline to execute on the appion
8 description computer.
  description It will create a few .mrcs files and the .star file. These files
10 description are necessary for Relion.

12 [FORM]
  Coordinates file; File; -d ; The downloaded coordinates from the appion website
14 Micrograph path; Directory; -m; Path to the original micrographs (/data/home/
    myamiimages/leginon..)
  Database ID; Text; -s ; The Database Name, visible with "ProjectDB -> processing
    db:" (ex ap21)
16 Session name; Text; -p ; The name of the experiment, visible with "ProjectDB ->
    Name within the List"
  Invert; Text; -i ; Invert the particle <b>(yes/no)</b>
18 Boxsize; Text; -n ; Boxsize for the particles in pixel
```

Abb. C.31: Beispiel einer Konfigurationsdatei für die GUI MDEMTols. Die Zeilen 1 bis 10 stellen den Kopfbereich mit den allgemeinen Programminformationen dar. Die Zeilen 12 bis 18 zeigen den Formularbereich mit 6 Eingabefeldern.

- (*tip*) Hinweis zur richtigen Auswahl
- (*command*) auszuführendes Kommando
- (*description*) Beschreibung des Programms

Abb. C.31 zeigt die Konfigurationsdatei von `Appion2Relion` und Abb. C.30 die entsprechende Repräsentation in der GUI. Um die jeweiligen Parameter (*category*, *name*, *tip*, *command*, *description*) zu definieren, werden die Stichwörter an den Anfang einer Zeile geschrieben, gefolgt von dem zuzuordnenden Wert. Der Parameter *category* kann hier mehrere durch Komma abgetrennte Kategorien enthalten. Der Parameter *tip* wird bei der Auswahl des jeweiligen Formulars innerhalb der GUI in der Statuszeile angezeigt (Abb. C.32) und soll bei der Auswahl des richtigen Tools helfen.

Der Parameter *description* gibt die Möglichkeit, eine Beschreibung bzw. eine Bedienungsanleitung anzuzeigen. Hierfür werden die einzelnen Zeilen zu einem Gesamttext zusammengefasst, neue Zeilen entsprechen einem Zeilenumbruch. Wei-

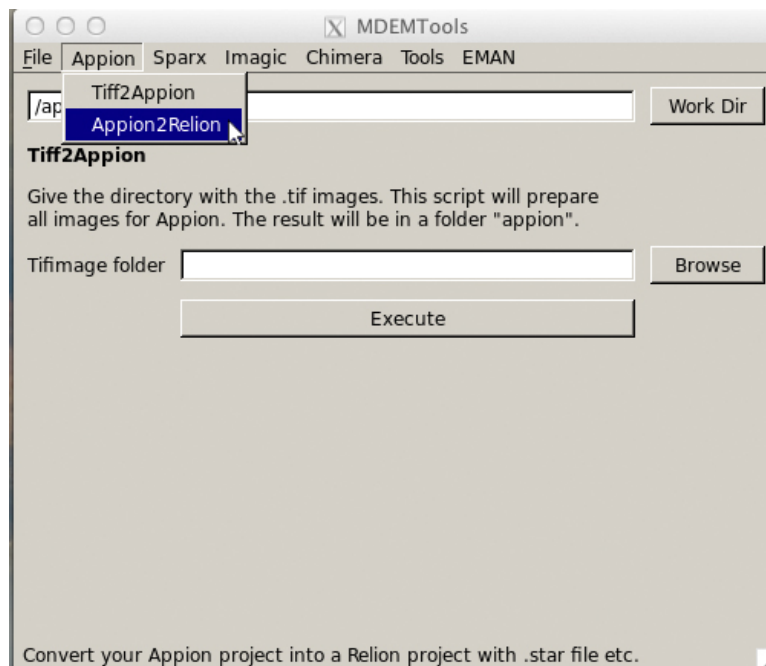
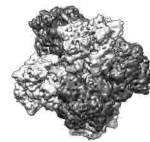


Abb. C.32: Grafische Benutzeroberfläche mit der in Abb. C.31 gezeigten Konfigurationsdatei vor der Auswahl von `Appion2Relion`. Die ausgewählte Kategorie `Appion` zeigt ein weiteres Werkzeug `Tiff2Appion`. Beim Halten der Maus auf einen Werkzeugnamen erscheint am Fuß des Programmfensters der Hinweistext (*tip*), hier „Convert your...“

terhin gibt es die Möglichkeit, mithilfe von Auszeichnungen den Text zu formatieren. Hier kann z. B. durch Einrahmen eines Textes in `text` dieser fett gedruckt (*bold*) dargestellt (siehe Abb. C.30) werden. Der Parameter *command* stellt den eigentlich Befehlsaufruf in der Kommandozeile dar. Auf diesen folgt die gleiche Anzahl an Platzhalter (*%s*), entsprechend der Menge der Eingabefelder im Formularbereich (im hier aufgeführten Beispiel sechs „%s“). Diese „%s“ werden durch die im Formular definierten Werte beim Ausführen ersetzt.

Der Formularbereich wird eingeleitet mit der Zeile `[FORM]`. Hier entspricht jede Zeile einem einzelnen Formularfeld. Die Angaben sind hier in vier Bereiche aufgeteilt, die durch ein Semikolon getrennt werden:

1. Beschriftung in der GUI
2. Art des Eingabefeldes (vier mögliche Werte: *Text*, *File*, *Directory*, *Save*)
3. Übergabeparameter für die Programmfunktion



4. ausführliche Beschreibung des Eingabefeldes

Dabei ergibt sich „Beschriftung“ und die „ausführliche Beschreibung“ im Vergleich der Konfigurationsdatei (siehe Abb. C.31 Zeile 17) mit der Abb. C.30 (vorletztes Eingabefeld). Die Art des Eingabefeldes wird abhängig vom zu übergebenden Wert gewählt, *Text* für frei gewählten Text oder Zahlen, *File* zur Auswahl einer vorhanden Datei, *Directory* zur Auswahl eines Verzeichnisses oder *Save* zur Auswahl des Speicherortes bzw. Datei. Hier kann eine bestehende Datei zum Überschreiben ausgewählt oder eine neue erzeugt werden. Abhängig vom Software-Tool wird eventuell ein Übergabeparameter benötigt, um den zu übergebenden Wert zuordnen zu können. Er besteht bei textbasierten Programmen häufig aus ein oder zwei Minus-Zeichen, gefolgt von kurzem Text oder auch nur einem Buchstaben. Dieser wird dann vor dem Ausführen vor den Wert gesetzt. Falls ein Textfeld nicht benutzt wird, wird auch kein Übergabeparameter ausgegeben.

4.3 Verwendete Programmierkonzepte

Bei der Entwicklung der GUI wurde auf aktuelle Programmierkonzepte Wert gelegt. Hier wurde mithilfe von testgetriebener Entwicklung (TDD) (Beck, 2002) sichergestellt, dass die GUI bei einer Erweiterung jederzeit auf Funktion getestet werden kann.

Die Versionierung wurde mit dem verteilten Versionskontrollsystem *Mercurial* (Mackall, 2006) vorgenommen. Dieses gestattet eine einfache Weitergabe an Entwickler und andere Arbeitsgruppen sowie das Erstellen von *software-patches*, um Fehler zu beseitigen (Loman and Watson, 2013). Dabei kann jede Änderung im Programmcode rückgängig gemacht werden, um z. B. einen funktionierenden Zustand wieder herzustellen.

Weiterhin wurde nach dem *Clean Code Codex* (Martin, 2008) programmiert. Durch diese Technik und Regeln wird ein gut lesbarer Programmcode erstellt, der auch für ungeübte Programmierer leicht zu verstehen ist.

4.4 Verbesserung der Administration am Rechen-Cluster

Auf dem Rechen-Cluster der Arbeitsgruppe wurde von einem einzigen, für alle Gruppenmitglieder angelegten Benutzer auf ein Mehrbenutzersystem umgestellt.

Daraus ergaben sich folgende Vorteile:

- Die verfügbare Speicherkapazität kann an jeden einzelnen Benutzer angepasst werden und der Festplattenspeicherplatz wird nicht überschritten.
- Durch eine Warteschlangenverwaltung (*queueing system*) kann die Rechenleistung effizient auf die Benutzer verteilt werden und eine Überlastung der Computer wird vermieden.
- Die Ordner sowie Programme werden mit spezifischen Benutzerrechten ausgestattet, z. B. nur lesen, lesen und schreiben, ausführen etc..

Neben dieser Umstellung wurde zudem ein inkrementelles *Backup* auf Basis von IBM Tivoli Backup Manager des kompletten hochleistungs Rechen-Clusters integriert. Hier werden die kompletten Daten der Benutzer täglich im Rechenzentrum der Johannes Gutenberg-Universität gespeichert und nach Löschung für weitere drei Monate archiviert. Ein Auszug der Datenbankinformationen von Appion wird als Textdateien (*database dump*) ebenfalls täglich mit gespeichert.

Die frei im Internet erhältlichen, von der Arbeitsgruppe verwendeten Programme wurden auf dem lokalen Rechen-Cluster installiert und so zur Verfügung gestellt. Bei Verfügbarkeit wurde während der Installation auf die Unterstützung von *message passing interface* (MPI) geachtet, um die Parallelisierung am Rechen-Cluster zu gewährleisten.

4.5 Dokumentation der neuen Software als DokuWiki

Auf dem Webserver des Rechen-Clusters wurde die Software DokuWiki installiert, zu erreichen unter „<http://zoo2cluster29.biologie.uni-mainz.de/wiki>“. Innerhalb der Arbeitsgruppen-Computer ist er ohne Passwort zugänglich, außerhalb wird ein Benutzername und Passwort benötigt.

Der DokuWiki ist organisiert in ein Hauptmenü auf der linken Seite, ein Inhaltsverzeichnis entsprechend des ausgewählten Wiki-Artikels auf der rechten Seite und den Kontent-Bereich in der Mitte (Abb. C.33). Das Hauptmenü ist unterteilt in die wichtigsten Themengebiete: Rechen-Cluster Administration, *HowTo*-Artikel und Tipps und Tricks zu installierten Programmen. Auf der rechten Seite



Tab. C.6: Hier ist eine unvollständige Liste an häufig verwendeten Markupbefehlen des DokuWiki.

Name	Version
"monospace"	Text zwischen zwei einzelnen Anführungszeichen wird als monospace Schrift gedruckt.
//kursiv//	Text zwischen zwei Schrägstrichen wird <i>kurisv</i> gedruckt.
fett	Text zwischen zwei Sternen wird fett gedruckt.
==== H1 =====	Text zwischen 6 Gleichzeichen wird zu einer Überschrift der Stufe 1.
===== H2 =====	Text zwischen 5 Gleichzeichen wird zu einer Überschrift der Stufe 2.
==== H3 =====	Text zwischen 4 Gleichzeichen wird zu einer Überschrift der Stufe 3.
<code>Code</code>	Text zwischen Code-Tags wird als spezieller Block für Skripte und Befehle dargestellt.
[[link optionaler Text]]	Link zu einer neuen Wiki Seite. Dabei wird <i>optionaler Text</i> als alternativer Text angezeigt.

befinden sich zusätzlich vier Schaltflächen: Zum Editieren des ausgewählten Artikels, anzeigen des Änderungsverlaufs, anzeigen der Verlinkungen zu diesem Wiki-Artikel und eine Schaltfläche „nach Oben“, um zum Anfang der Seite zu springen. Im Änderungsverlauf können alle Versionen eines Wiki-Artikels wieder hergestellt werden. Hier sind die Änderungen einzusehen und der Autor wird gespeichert. Abb. C.33 zeigt exemplarisch den Wiki-Artikel für die MD-Simulationssoftware **Gromacs**.

Die Wiki-Software ist so konfiguriert, dass die Seiten von jedem Benutzer editiert werden können. Durch die Mitarbeiter in der Arbeitsgruppe werden somit Texte des Autors um Details erweitert. Der DokuWiki wurde neuen Arbeitsgruppenmitgliedern zur Verfügung gestellt, die bei Unklarheiten den Autor des betreffenden Wikis fragen können. Dieser kann nach Bedarf die Texte anpassen und Ergänzungen einfügen. Durch den Wiki ist ein relativ selbstständiges Einarbeiten neuer Mitarbeiter möglich.

Zum Editieren der Texte wird innerhalb eines Webbrowsers ein Editorfenster zur Verfügung gestellt, dass die wenigen nötigen Auszeichnungen (Überschriften, Verlinkungen, Abbildungen) bei Bedarf einfügt. Beispiele für eine Überschrift oder kursiven Text sind in Tabelle C.6 aufgeführt. Neben dieser Art der Formatierung stehen auch Möglichkeiten zur Verfügung, um Bilder in die Texte einzubauen.



Hauptmenü

- Start
- Skripte und Helferlein
- Administration des Clusters
- 29/2ser Administration
- NFS-SRV Administration
- Temperaturfühler
- Arbeitsplatzrechner
- HowTos
 - 3D-EM Workflow
 - Scanner HowTo
 - 3D Drucker HowTo
 - Drucker HowTo
- Einzlersammlung
- Programme
- Terminal Befehle
- CTF
- Videos erstellen, konvertieren, etc.
- Imageic
- Frealign
- EWMAN
- EMAN2
- IMOD
- Unsere Queue
- Amira
- Chimera
- Proteome
- SPARKX
- Appion
- DireX
- Gromacs
- VMD
- Scripteller Archiv
- Dokumente (HowTo, etc)
- Generous Administration
- Vacuum Citing
- alexTIPS und Tricks

Um mit Gromacs arbeiten zu können muss man erst mal den folgenden Exportbefehl in seine `.bashrc` eintragen, und danach `source`.

```
echo "export PATH=/apps/prod/gromacs/bin:$PATH" >> ~/.bashrc
source ~/.bashrc
```

Weiterhin gilt, alle Parameter die zwischen `<>` angegeben sind durch eigene zu ersetzen! Hier werden nur teilweise sinnvolle Namen benutzt.

Vorbereitungen

Gromacs rechnet sehr lange. Als Beispiel: Zwei Monomere von Limulus (ca 1300 Aminosäuren) rechnen rund 12h für 4ns Simulation auf einem beo. 4 Untereinheiten rechnen rund 30h für 4ns. Man sieht also schon, dass die Zeit eher exponentiell zunimmt.

Zur Zeit funktioniert das parallele Rechnen auf mehreren Knoten leider nicht. Es wird empfohlen Gromacs auf einem der neuen Knoten zu starten, da werden dann 32 Kerne parallel benutzt. Schaut hierzu bitte in das Queue File am Ende.

Die normale Vorgehensweise sieht in etwa so aus (die wichtige Ausgabedateien sind dick angegeben):

- Protein für Gromacs vorbereiten **<01_protein.pdb>**
- Protein in eine noch leere Box packen **<02_protein-box.pdb>**
- Box mit Wasser auffüllen **<03_protein-with-solvent.pdb>**
- Energiegleich minimieren **<04_em-file.tpr>** **<05_protein-minimiert.pdb>**
- Wassermoleküle mit Ionen austauschen **<06_em-file.tpr>** **<07_protein-ion.pdb>**
- Energiegleich minimieren **<08_em-file.tpr>** **<09_protein-ion-minimiert.pdb>**
- Jetzt erst mal das Wasser equilibrieren **<10_eq-water.tpr>** **<11_minimiertes-wasser.pdb>**
- Eigentliche Molecular Dynamics Simulation **<12_md-file.tpr>** **<13_ENDERGEBNIS.pdb>**

Table of Contents

- Vorbereitungen
- PDB Datei vorbereiten
- Den korrekten Protolienungszustand errechnen
- Die pdb vorbereiten
- Protein in eine noch leere Box packen
- Box mit Wasser auffüllen
- Minimieren der Struktur
- Minimierung vorbereiten
- Box mit Wasser füllen
- Erneut minimieren mit Wasser
- Starten der eigentlich MD
- Visualisierung mit VMD
- Videos
- Beispiel Queue Datei
- MDP Parameter erklärt
- Miss noch eingepflegt werden:
- Liganden
- Skripte die helfen könnten
- PDB Zwischenstände zu Trajektorie zusammenschaffen
- Abmessungen mit `g_dist`
- Atome in Reichweite
- Diagramme
- Trajektkoren in Chimera

Edit

Abb. C.33: DokuWiki auf dem Rechen-Cluster der Arbeitsgruppe, am Beispiel des Gromacs-Artikels. Im grünen Rahmen sieht man das Hauptmenü und in rot eingerahmt das Inhaltsverzeichnis für diesen Artikel. Im Blauen Rahmen sind die Bedienelemente zum Editieren, Änderungsverlauf und Verlinkungen auf die aktuelle Seite.



4.6 Entwickelte Werkzeuge

Die von mir entwickelten Software-Tools sind alle konsolenbasierend und in verschiedenen Programmiersprachen entwickelt. Eine genauere Bedienungsanleitung findet sich im DokuWiki. Die meisten Programme geben eine kurze Anleitung aus, wenn sie ohne Übergabeparameter aufgerufen werden. Wenn sie in die GUI aufgenommen wurden, wird die entsprechende Kategorie hier mit angegeben.

4.6.1 Software-Tools zum Arbeiten mit MD-Simulationen

gromacsgraph.py Um die gemessenen Abstände in einer Grafik aufzutragen, wurde das Programm `gromacsgraph.py` entwickelt. Es liest beliebig viele Dateien mit Abmessungen ein und stellt eine Grafik mit Legende der Dateinamen dar. Es wurde in der Programmiersprache `Python` geschrieben und ist nicht in der GUI verfügbar.

gromacshist.py Um die gemessenen Abstände in einem Balkendiagramm aufzutragen, wurde das Programm `gromacshist.py` entwickelt. Hierbei werden beliebig viele Dateien mit Abmessungen zu einem Diagramm zusammengefasst. Die Abstände werden dabei gruppiert auf ganze Ångström und alle Werte über 15 Å werden in einer Gruppe zusammengefasst. Das Skript wurde in der Programmiersprache `Python` geschrieben und ist nicht in der GUI verfügbar.

gromacsdisc.py Um die gemessenen Abstände in einem Scheibendiagramm darzustellen, wurde das Programm `gromacshist.py` entwickelt. Es liest beliebig viele Dateien mit Abmessungen ein und stellt für jede Datei ein eigenes Diagramm. Es wurde in der Programmiersprache `Python` geschrieben und ist nicht in der GUI verfügbar.

restrain35angst.sh Um bestimmte Atome im Raum zu fixieren, wurde das Programm `restrain35angst.sh` entwickelt. Es benötigt eine Koordinate im 3D-Raum und schreibt danach alle Atome außerhalb eines 35 Å-Radius in eine Datei. Diese Datei wird danach in `Gromacs` verarbeitet. Das Programm wurde in der Programmiersprache `AWK` geschrieben und ist nicht in der GUI verfügbar.

```
oneletter = {'ALA': 'A', 'ARG': 'R', 'ASN': 'N', 'ASP': 'D', 'CYS': 'C', 'GLN': 'Q',  
2   'GLU': 'E', 'GLY': 'G', 'HIS': 'H', 'ILE': 'I', 'LEU': 'L', 'LYS': 'K', 'MET': 'M',  
   'PHE': 'F', 'PRO': 'P', 'SER': 'S', 'THR': 'T', 'TRP': 'W', 'TYR': 'Y', 'VAL': 'V'}  
4 for i in chimera.selection.currentResidues():  
   print "[%s%s_%s]" % (oneletter.get(i.type, i.type),  
6       i.id.position,  
       i.id.chainId)  
8   for j in chimera.selection.currentAtoms():  
     if j.residue == i:  
10      print j.serialNumber,  
      print ""
```

Abb. C.34: Programmcode zur Erstellung der Atomgruppen innerhalb von Chimera. Nach dem Ausführen wurden die Gruppen mit den zugehörigen Atomnummern in IDLE ausgegeben (Abb. C.21).

md2volume.sh Um aus den Daten der MD-Simulationen eine 3D-Dichtekarte zu erstellen, wurde das Programm `md2volume.sh` entwickelt. Es extrahiert zu bestimmten Zeitpunkten PDB-Dateien aus der MD-Simulation und simuliert daraus einzelne 3D-Dichtekarten. Im Anschluss werden alle Einzolvolumina zu einem Gesamtvolumen addiert. Das Programm wurde mit `bash`-Befehlen programmiert und ist nicht in der GUI verfügbar.

Auswahl der Atome zum Vermessen Um die Atome zum Abmessen auszuwählen, müssen die Atomnummern in eine Datei geschrieben werden. Hierfür wurde ein Script entwickelt, das mithilfe des Programms Chimera eine Liste der selektierten Atome mit passenden Gruppennummern generiert. Hierfür muss in Chimera die IDLE-Programmumgebung geöffnet und der Programmcode aus Abb. C.34 eingegeben werden.

4.6.2 Software-Tools zum flexiblen Fitten

hbond2direx Die Datei mit den Wasserstoffbrücken-Informationen wird im Programm Chimera erstellt. Diese wird durch `hbond2direx` in ein spezielles Format konvertiert. Es ist in der Sprache AWK programmiert und in der Kategorie *Direx* innerhalb der GUI verfügbar.



4.6.3 Software-Tools zum Erstellen und Bearbeiten von 3D-Rekonstruktionen

reduceAngles Bei negativkontrastierten 3D-Rekonstruktionen kann es häufig zu Vorzugsansichten kommen, da sich die Proteine in bevorzugter Orientierung auf dem Kohlefilm anheften. Hier kann es sinnvoll sein, bei der Erstellung der Referenzen auf bestimmte Winkel zu verzichten. Das Programm `reduceAngles` reduziert die Referenzen für die Programme `IMAGIC` und `EMAN/EMAN2`. Es ist in der Programmiersprache `AWK` programmiert und in den Kategorien `IMAGIC` und `EMAN` in der GUI verfügbar.

appion2relicion `Relion` (Scheres, 2012) ist eine 3D-Rekonstruktionssoftware, die in dieser Arbeit nicht zum Einsatz kam, jedoch in der Arbeitsgruppe verwendet wird. Um die in `Appion` bearbeiteten Datensätze in `Relion` benutzen zu können, wurde ein Programm entwickelt, das die benötigten Daten aus der `Appion`-Datenbank sammelt und entsprechend konvertiert. Das Programm wurde in der Programmiersprache `Python` geschrieben und ist in der Kategorie `Relion` innerhalb der GUI verfügbar.

anglespread2.py Zur Analyse von bevorzugten Ansichten wurde das Programm `anglespread2.py` entwickelt. Es wurde in der Programmiersprache `Python` geschrieben und benötigt eine Datei mit den Winkeln der einzelnen Partikel. Daraus erstellt es eine `.bild`-Datei, die mit der entsprechenden 3D-Rekonstruktion in `Chimera` geöffnet wird. Hier werden nun Zylinder um die 3D-Rekonstruktion angezeigt, deren Länge die Anzahl der Partikel mit dieser Orientierung widerspiegelt (Abb. C.35). Das Programm ist in den Kategorien `IMAGIC`, `EMAN` und `Sparx` in der GUI verfügbar.

ImagicRefinement.b Die Software `IMAGIC` stellt kein Modul zum automatischen Verfeinern von Strukturen zur Verfügung. `ImagicRefinement.b` ist ein `bash`-Programm, das `IMAGIC` ähnliche Funktionen wie `Sparx` verleiht, z. B. eine bestimmte Anzahl von Iterationen Verfeinern und schrittweise die Winkel der Referenzen verkleinern. Das Programm ist in der Kategorie `IMAGIC` in der GUI verfügbar.

tif2appion.b Um die EM-Aufnahmen der TemCam-F416 Digitalkamera vom TIFF-Bildformat in ein für `Appion` lesbares Format umzuwandeln, wurde das

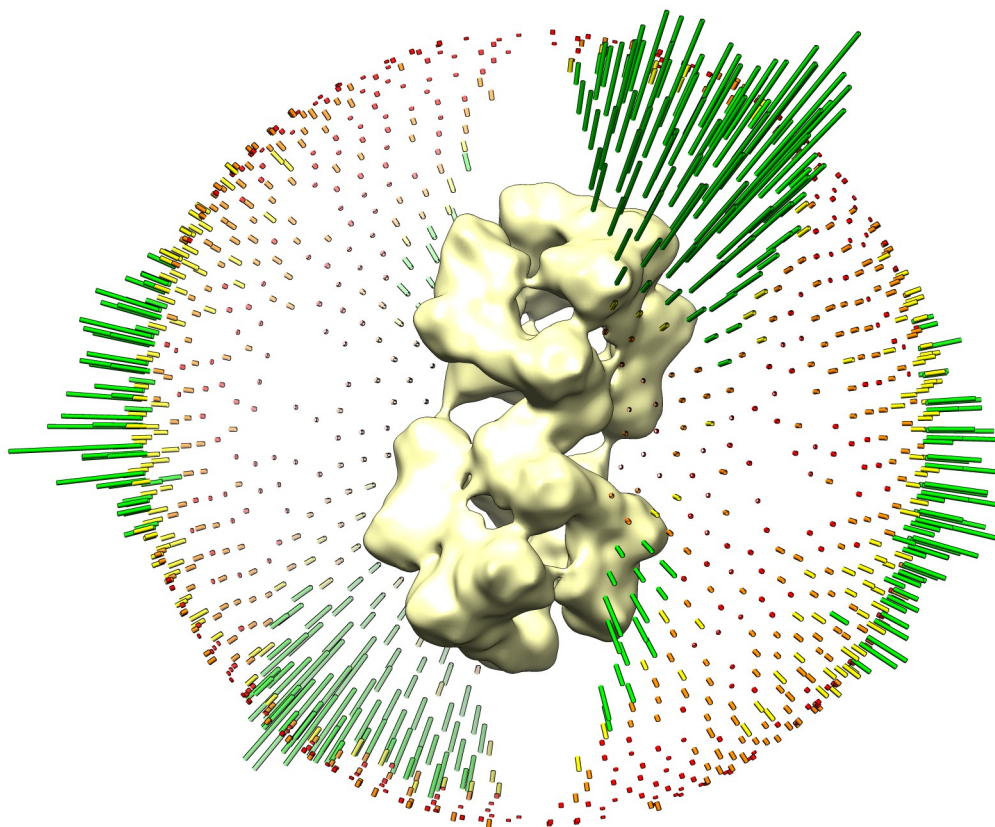


Abb. C.35: Das Programm `anglespread2.py` erstellt eine `.bild`-Datei zum Öffnen in Chimera. Zusammen mit der 3D-Rekonstruktion kann die Verteilung der Partikelansichten analysiert werden. Jeder Zylinder entspricht einer Orientierung im 3D-Raum, die Länge der Zylinder spiegelt die Anzahl der Partikel mit dieser Orientierung wider.

Programm `tif2appion.b` entwickelt. Es konvertiert automatisch alle `.tif`-Dateien und verschiebt sie in einen Ordner `appion` im aktuellen Verzeichnis. Dieser Ordner wird beim Datenimport in `Appion` angegeben. Das Programm ist ein `bash`-Programm und in der Kategorie *Tools* in der GUI verfügbar.

sparxres.sh Das Programm `sparxres.sh` sucht nach allen Dateien im aktuellen Ordner und gibt die Auflösung in der Kommandozeile aus, ohne eine FSC-Grafik zu erzeugen. Es ist ein `bash`-Programm und steht nicht in der GUI zur Verfügung.



plotfsc.py Um die FSC-Grafik zu erstellen und die Auflösung zu bestimmen, wurde das Programm `plotfsc.py` erstellt. Es ist in der Programmiersprache Python geschrieben und ist in der GUI in der Kategorie *Tools* verfügbar.

4.6.4 Allgemeine Software-Tools

readTifHead.py Um die TemCam-F416 Digitalkamera spezifischen Daten aus einer EM-Aufnahme auszulesen wurde das Programm `readTifHead.py` in der Programmiersprache Python entwickelt. Es gibt Informationen wie die Belichtungszeit, die eingestellte Vergrößerung und die Position des Probenhalters für das ausgewählte Bild aus. Diese Informationen sind sonst nur direkt am Elektronenmikroskop zu erfahren. Das Programm ist in der Kategorie *Tools* in der GUI verfügbar.

16to8bittif Die von der TemCam-F416 Digitalkamera erstellten Aufnahmen sind in einem 16bit TIFF-Format, das nur von Spezialprogrammen lesbar ist. Das Programm `16to8bittif` konvertiert die Bilder automatisch in ein Standardbildformat. Es ist ein `bash`-Programm und in der GUI in der Kategorie *Tools* verfügbar.



D Diskussion

1 3D-EM des multimeren Hämoglobins von *Daphnia pulex*

Für das *D. pulex*-Hämoglobin galt es, die Anzahl der Häm-Domänen zu bestimmen da die bis dahin angewendeten Methoden zur Bestimmung der Gesamtmasse des Proteins zu keinem eindeutigen Ergebnis gekommen waren. In der Ultrazentrifuge war eine Proteinmasse von 420 kDa bestimmt worden (Tabelle D.1) (Svedberg and Eriksson-Quensel, 1934), über Gelfiltrationschromatographie eine Masse zwischen 417 und 470 kDa (Dangott and Terwilliger, 1980; Peeters et al., 1990).

Aus der Literatur war bekannt, dass die Untereinheit dieses Hämoglobins aus zwei Häm-Domänen besteht (Dangott and Terwilliger, 1980; Peeters et al., 1990). Auch die Sequenz des *D. pulex*-Hämoglobins wurde ermittelt, wobei elf Hämoglobingene identifiziert wurden (Colbourne et al., 2011). Des Weiteren war bekannt, dass sich zwei Untereinheiten über Disulfidbrücken zu einem Dimer zusammenschließen können, welches folglich dann vier Häm-Domänen enthält (Peeters et al.,

Tab. D.1: Massenbestimmung bei *Daphnia*-Hämoglobinen. Die Tabelle führt die bestimmten Massen sowie die verwendete Methode für den jeweiligen Organismus auf.

Autor	Organismus	Masse in kDa	verwendete Methode
Peeters et al. (1990)	<i>D. pulex</i>	417-450	Gelfiltrations-Chromatographie
Dangott and Terwilliger (1980)	<i>D. pulex</i>	430-470	Gelfiltrations-Chromatographie
Svedberg and Eriksson-Quensel (1934)	<i>D. pulex</i>	420	Sedimentationskonstante
Lamkemeyer et al. (2005)	<i>D. magna</i>	580	16x36kDa (Seq. bekannt)
Ilan et al. (1982)	<i>D. magna</i>	505± 35 & 483± 27	Sedimentationsgleichgewicht
Sugano and Hoshi (1971)	<i>D. magna</i>	670	Sedimentationskonstante

1990; Dewilde et al., 1999). Aufgrund der bis zu diesem Zeitpunkt bekannten Molekularmassen und der bestimmten Masse von ca. 36 kDa pro Untereinheit wurde deren Anzahl in der Quartärstruktur auf 12 geschätzt (Svedberg and Eriksson-Quensel, 1934; Dangott and Terwilliger, 1980; Peeters et al., 1990; Gerke et al., 2011; Zeis et al., 2013). Demnach sollte das *D. pulex*-Hämoglobin aus 24 Häm-Domänen bestehen.

Die von mir vorgestellte 14 Å-Rekonstruktion des Hämoglobins von *D. pulex* zeigt 32 Massen, die jeweils einer Häm-Domäne entsprechen. Das ergibt zusammen ca. 580 kDa. Die publizierten Molekularmassen sind also offenbar falsch. Für eine genauere Bestimmung, wie die Häm-Domänen räumlich orientiert sind bzw. wie sie sich zu Dimeren aneinander lagern, war die Auflösung nicht ausreichend. Aufgrund der Ergebnisse gehe ich davon aus, dass es sich um 32 Häm-Domänen handelt bzw. dass das Protein aus 16 Untereinheiten besteht.

Für die verwandte Art *Daphnia magna* wurden in der Vergangenheit ebenfalls unterschiedliche Molekularmassen des Hämoglobins publiziert (siehe Tabelle D.1). Abhängig von der verwendeten Methode wurden Werte zwischen 483 und 670 kDa ermittelt. Diese Differenz von 187 kDa entspricht ca. 11 Häm-Domänen. Anhand der Sequenz wurden für die Quartärstruktur, wie von mir bei *D. pulex*, 16 Polypeptidketten mit je 36 kDa angenommen (Lamkemeyer et al., 2005), was 580 kDa entspricht. Direkte Vergleiche von EM-Aufnahmen (Abb. C.1 Seite 56) lassen vermuten, dass die Hämoglobine von *D. pulex* und *D. magna* vergleichbar groß und ähnlich strukturiert sind. Zudem haben die Hämoglobin-Gene eine Sequenzidentität zwischen 70 und 73% (Colbourne et al., 2011; Gerke et al., 2011).

Warum sich die beiden Hämoglobine bei Molekulargewichtsbestimmungen so unterschiedlich verhalten ist unklar. Am naheliegensten wären Zuckerreste am Hämoglobin. Diese konnten für *D. magna* nachgewiesen werden. Hier wurden O-verkettete Zucker gefunden (Lamkemeyer et al., 2006). Bei *D. pulex* wurden N-verkettete Zucker ausgeschlossen (Peeters et al., 1990). Durch die Sequenzanalyse wurden mithilfe von **NetOGlyc Server 3.1** (Julenius et al., 2005) auch potentielle Zuckerbindestellen am N-terminalen Linker bestimmt. Durch Deglykosylierung konnten aber keine Unterschiede in der Masse der Hämoglobinuntereinheiten nachgewiesen werden (Gerke et al., 2011). Hier traten wohl nur leichte Unterschiede im isoelektrischen Punkt auf (Gerke et al., 2011). Dass die Unterschiede bei der Molekulargewichtsbestimmung von Zuckern herrühren, ist daher



unwahrscheinlich.

Vielleicht unterscheiden sich beide Hämoglobine in ihrer Konformation, was andere Laufeigenschaften in der Gelfiltrations-Chromatographie ergeben könnte. Hierfür wurden die bekannten Proteinsequenzen von *D. pulex*-Hämoglobin untersucht und es wurde in einer der Sequenzen (Abb. D.2) ein zweites Cystein gefunden (UniProt ID E9GCI5_DAPPU (Colbourne et al., 2011)). Da für die aus zwei Häm-Domänen bestehende Untereinheit keine Kristallstruktur vorliegt, wurde von mir unter Verwendung des Programms **Chimera** und des Moduls **Modeller** (Eswar et al., 2007; Yang et al., 2012b) eine Struktur homologiemodelliert (Abb. D.1). PDBid 1X46 (Kuwada et al., 2007) passt von der Sequenz gut zur N-terminalen Häm-Domäne, PDBid 1HLM (McDonald et al., 1992) zur C-terminalen Domäne. Dazwischen befindet sich ein 15 Aminosäuren langer Linker sowie N-terminal ein 50 Aminosäuren langer Bereich. Aus Abb. D.1 wird deutlich, dass sich die Häm-Domänen aufgrund des langen Linkers fast beliebig zueinander orientieren können. Dazu wird eine Kristallstruktur benötigt, um die Anordnung zu bestimmen.

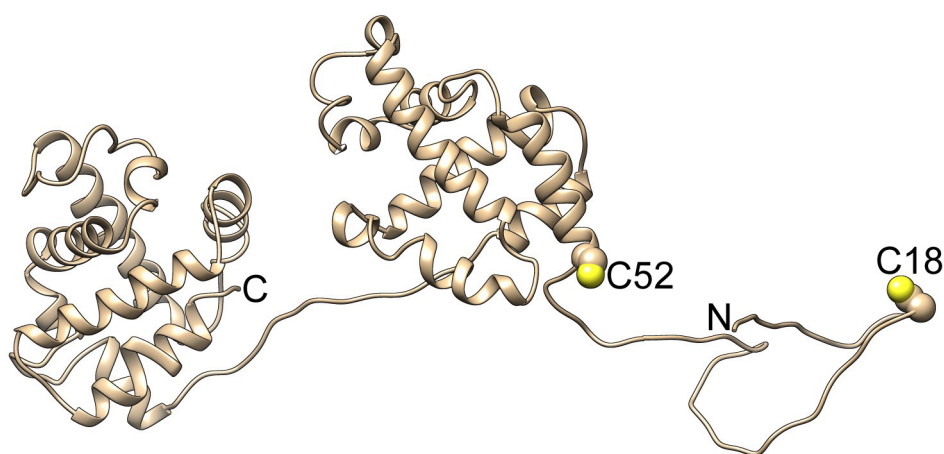


Abb. D.1: Homologiemodell einer Untereinheit von *D. pulex* (Templates: PDBid 1X46 und 1HLM). Die Häm-Domänen sind über 15 Aminosäuren verbunden, der N-terminale Linker ist 50 Aminosäuren lang.

```
>tr|E9GCI5|E9GCI5_DAPPU Hemoglobin OS=Daphnia pulex GN=DAPPUDRAFT_316317 PE=3
MLASFklvllsvvalaCAWPQFGSSMTTGPTTSTVPAKENSQGTPKLDCFDHSIIRN
TWDQAKKNGEFAPQVLLRFIKAHPEYQQMFGKFASVPHYNLLRNGDFLAQAYTISAGLNV
VIQSLSSQELLAAQFNLLGSAYQPRGVTPAMFEEFSVILEQVLEETLGSTFNVEARKAWN
KGMVAIIAGISKTLKNPEDLADAQSNLTRPQIRNVQRSWESMKSGRNSLVSAIFIKLFKE
TPRVQKHFAKFANVPVDSLGRNGDYIQQVALVADRLDTLISAMDDQLQLLGNINYLKYTH
AKRSIPRKTWEDFARLLVELLPTRGVSASDVESWKGVTTLVNGIAPKN
```

Abb. D.2: *D. pulex*-Hämoglobin Sequenz einer Untereinheit. Die zwei Cysteine sind gelb hervorgehoben.

Die beiden Cysteine kommen jedoch nicht bei allen Sequenzen von *D. pulex*-Hämoglobin vor und es ist unklar, aus welchen der bekannten Sequenzen das native Hämoglobin aufgebaut ist. Von den 14 Hämoglobin-Genen besitzen nur elf die volle Länge zur Kodierung der 2-Domänen-Hämoglobinkette (Gerke et al., 2011). Bei zwei der entsprechend langen Sequenzen konnte kein N-terminales Signalpeptid gefunden werden und sie könnten somit nicht über Vesikel in die Hämolymphe ausgeschleust werden (Gerke et al., 2011). Jedoch entspricht die veröffentlichte Sequenz (UniProt ID E9GCI5_DAPPU) keiner der in Gerke et al. (2011) gezeigten Sequenzen, somit ist nicht klar, ob das zweite Cystein die unterschiedlichen Laufeigenschaften bei der Massenbestimmung verursachen könnte.

Es gab Hinweise, dass sich in der Quartärstruktur zwei Untereinheiten über Disulfidbrücken zu einer kooperativen, funktionellen Einheit zusammensetzen. Dieses Dimer mit vier Häm-Domänen wäre dann ein stabiler Baustein, aus dem sich die Quartärstruktur des Proteins aufbaut (Dewilde et al., 1999). Ein Threoninreicher Bereich am N-Terminus soll als verlängerter Linker dienen und die Bildung von Tetrameren vereinfachen (Dewilde et al., 1999). Falls Tetramere (acht Häm Domänen) vorliegen, dann wären auch nur Quartärstrukturen mit einem Vielfachen von vier möglich. Aufgrund der 4-fach Symmetrie, die man an den Klassensummen sehr gut erkennen kann (Abb. C.1 Seite 56), spricht dieses ebenfalls für $4 \times 4 = 16$ Untereinheiten (32 Häm-Domänen).

Beim Hämoglobin des Crustaceen *Triops cancriformis* wurden ähnliche Probleme bei der Massenbestimmung beschrieben. Hier konnte die Masse mit einer Kombination aus nativer Gelelektrophorese und Elektronenspray-Ionisations Massenspektrometrie (ESI-MS) genau bestimmt werden (Rousselot et al., 2006).



2 3D-EM des 2x6meren Hämocyanins von *Astacus leptodactylus*

Bei *A. leptodactylus*-Hämocyanin sollte die Orientierung der beiden Hexamere zueinander geklärt und die Bindestelle(n) zwischen den beiden Hexameren identifiziert werden. Die erreichte Auflösung von 15 Å war ausreichend, um das kristallographische Hexamer von *Panulirus interruptus*-Hämocyanin (PDBid: 1HCY; (Volbeda and Hol, 1989)) als starren Körper stabil einzupassen. Hierbei stellte sich heraus, dass es nur eine deutliche Kontaktstelle gibt. Dort nähern sich beide Hexamere mit je einer Untereinheit bis auf 4 Å an, und zwar jeweils mit dem 7-strängigen β -Barrel von Domäne 3.

Um genauere Aussagen über die Kontaktstelle treffen zu können, müsste die Auflösung durch einen größeren Datensatz verbessert werden. Weiterhelfen würde auch ein Homologiemodell der am Kontakt beteiligten Hämocyanin-Untereinheit. Von *A. leptodactylus* ist bekannt, dass die beiden Hexamere mit einer Disulfidbrücke verbunden sind (Stöcker et al., 1988). Ein topographisches Modell wurde von Stöcker et al. (1988) erstellt, das mit meiner 3D-Rekonstruktion des Hämocyanins formal vergleichbar ist. Aufgrund der Näherung der Domäne 3 bis auf 4 Å liegt die Vermutung nahe, dass an dieser Stelle ein ungepaartes Cystein vorliegen muss und somit eine Disulfidbrücke zum zweiten Hexamer aufbauen kann.

Weiterhin wurde in Stöcker et al. (1988) von der typischen Rechteck-Hexagon Ansicht gesprochen, die darauf zurückgeführt wurde, dass die Hexamere um ca. 90° gegeneinander verdreht sind. Abmessungen an der 3D-Rekonstruktion des *A. leptodactylus*-Hämocyanins ergeben eine Verdrehung von 103° gegen den Uhrzeigersinn (Abb. D.3 (A) bis (C)). Die Zentren der beiden Hexamere sind 110 Å voneinander entfernt und die Gesamtlänge des Proteins beträgt 240 Å. Das obere Hexamer ist im Vergleich zum unteren um 55° im Uhrzeigersinn um die Rotationssymmetrieachse verdreht. Die Werte sind in Tabelle D.2 zusammengefasst und den Abmessung eines 2x6mers aus der 7,1 Å-Rekonstruktion des *L. polyphemus*-Hämocyanins gegenüber gestellt. Zudem sind die Abmessungen eines 2x6meren Hämocyanins der Strandkrabbe *Carcinus aestuarii* (Mičetić et al., 2010) aufgeführt. An diesem Hämocyanin wurden die Abmessungen mithilfe von Kleinwinkelstreuung (*Small-angle X-ray scattering*) (SAXS) bestimmt. Der Wert für die Drehung um die Rotationssymmetrieachse war in dieser Publikation nicht

Tab. D.2: Abmessung unterschiedlicher 2x6mere von *Astacus leptodactylus*, *Limulus polyphemus* und *Carcinus aestuarii* (Mičetić et al., 2010).

	<i>A. leptodactylus</i>	<i>L. polyphemus</i>	<i>C. aestuarii</i>
Höhe	240 Å	210 Å	225 Å
Breite	145 Å	140 Å	—
Entfernung der Zentren	110 Å	99 Å	106 Å
Drehung der Hexamere gegeneinander	103°	105°	125°
Kippung der Rotationssymmetrieachse	3°	3°	—
Drehung um Rotationssymmetrieachse	55°	3°	15–20°*

* Der Winkel wurde anhand der Abbildung (Fig. 6 A) in Mičetić et al. (2010) geschätzt.

angegeben und wurde anhand der Abbildungen geschätzt. Bei höheren Auflösungen könnten sich die Maße noch etwas ändern (vergleiche Kapitel D.3.4).

Als Fazit ergibt sich, dass nach derzeitigem Stand bei den drei Hämocyaninen die Drehwinkel der beiden Hexamere signifikant unterschiedlich sind. Die Dimerisierung von Hexameren hat demnach konvergent stattgefunden.

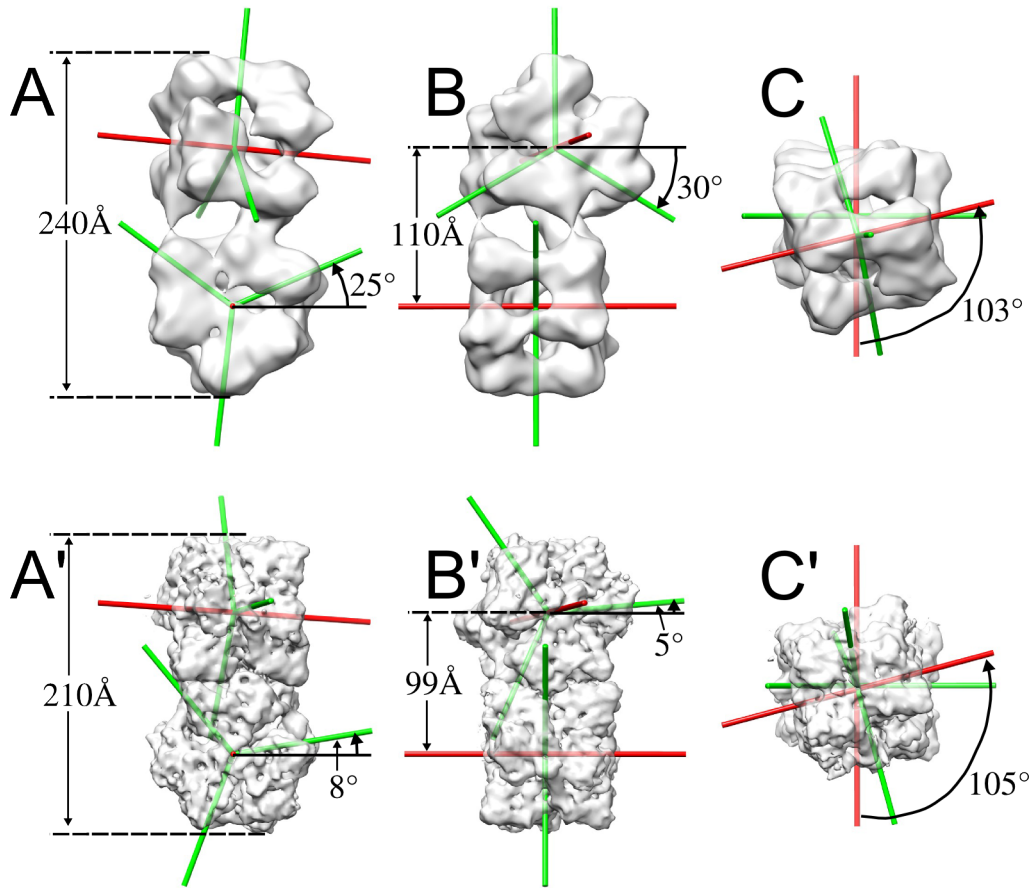
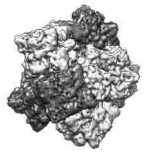


Abb. D.3: Vergleich des *A. leptodactylus*-Hämocyanins und extrahierten 2x6mer aus der 7,1 Å-Rekonstruktion des *L. polyphemus*-Hämocyanins. (A) Frontansicht des *A. leptodactylus*-Hämocyanins und Drehung um die Rotationssymmetrieachse des unteren Hexamers sowie dessen Länge. (B) Seitenansicht des *A. leptodactylus*-Hämocyanin und Drehung um die Rotationssymmetrieachse des oberen Hexamers sowie die Entfernung der Zentren. (C) Aufsicht des *A. leptodactylus*-Hämocyanins mit eingezeichneter Drehung der Hexamere gegeneinander. A' bis C' zeigen zum Vergleich *L. polyphemus* 2x6mer mit gleichen Abmessungen.

3 3D-EM des 8x6meren Hämocyanins von *Limulus polyphemus*

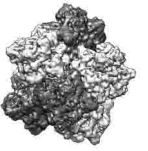
Mit der erreichten Auflösung von 7,1 Å war die 3D-Rekonstruktion des 8x6meren Hämocyanins von *L. polyphemus* deutlich besser als die publizierte 12 Å-Rekonstruktion von Martin et al. (2007). Dadurch konnten die Brücken zwischen den acht Hexameren mit größerer Genauigkeit bestimmt werden und das publizierte pseudoatomare Modell wurde durch flexibles *Fitting* verfeinert. Um exakte Informationen über die Bindungen zwischen den Hexameren zu erhalten, wurden exemplarisch an einer der Kontaktstellen (dem IIIA–IIIA-Interface) MD-Simulationen durchgeführt.

3.1 MD-Simulation des *L. polyphemus*-Hämocyanin

Um die Methode der MD-Simulation in der Arbeitsgruppe zu etablieren, wurden einfache Umgebungsbedingungen gewählt, die schrittweise um Details erweitert werden sollten. Somit konnte überprüft werden, ob die Rechnungen aufgrund der hohen Rechenzeiten überhaupt auf dem lokalen Rechen-Cluster durchgeführt werden können. Das erste Umgebungsmedium (Puffer A) bestand aus H₂O und 10 mM NaCl. Im zweiten Versuch wurde zusätzlich 10 mM MgCl₂ hinzugefügt (Puffer B). In Zukunft ist geplant, Trishydroxymethyl-aminomethan (TRIS) als Puffermolekül hinzuzufügen. Somit könnte man den *in vitro* verwendeten Puffer nachbilden.

Die MD-Simulationen lieferten interessante Ergebnisse zur Interaktion zwischen den Hexameren an der IIIA–IIIA-Kontaktstelle, aber die Rechenzeit war sehr hoch. Eine 100 ns Simulation der vier Untereinheiten dauerte auf einem Rechenknoten des Rechen-Clusters etwa 35 Tage. Die Nutzung des kompletten Rechen-Clusters wurde von der derzeitigen Installation (noch) nicht unterstützt. Hier muss zukünftig Unterstützung bei den Entwicklern eingeholt werden, um die vollständige Parallelisierung auf dem Rechen-Cluster zu ermöglichen.

Als Indiz für eine Bindung wurde nur der räumliche Abstand zwischen den Aminosäureresten beachtet und nicht die Bindungsstärke der Salzbrücken. Diese könnte in künftige Berechnungen einbezogen werden. Da sich die Untereinheiten aufgrund der Fixierung nicht voneinander entfernen konnten, war für meine



Versuche die Bindungsstärke vernachlässigbar.

Im direkten Vergleich mit Puffer A (Abb. D.4) fällt auf, dass ein Magnesium Ion (Mg^{2+}) an die Aminosäuren IIIA₁D136 und IIIA₁D422 bindet. Die ebenfalls in der Abb. angezeigte Aminosäure IIIA₂D136 ist ca. 7 Å entfernt. Hier besteht die Annahme, dass die Verbindung IIIA₁H134 ↔ IIIA₂D514 die Bindung des Mg^{2+} -Ions begünstigt, da es den *Loop* mit der Aminosäure IIIA₁H134 in Richtung IIIA₂ zieht (Abb. D.4 B). Theoretisch ist aufgrund der Symmetrie für die IIIA₂-Untereinheit ein ähnliches Verhalten zu erwarten und muss in mehreren Simulationen untersucht werden.

3.2 Fixierung von Aminosäuren vor der MD-Simulation

Durch die 7,1 Å-Rekonstruktion waren Position und Abstände der vier betrachteten Untereinheiten gut definiert. Für erste Tests wurde der Bereich außerhalb von 35 Å um die Kontaktstelle fixiert. Somit konnte sich nur die Kontaktstelle frei bewegen. In Zukunft muss hier mehr Flexibilität erlaubt werden, um die Bewegung der kompletten Untereinheit zu erfassen. Hierfür sollen nur die äußersten Teile des pseudoatomaren Modells fixiert werden (Abb. D.5). Durch diese erweiterte Flexibilität wäre es z. B. möglich, Änderungen im aktiven Zentrum zu sehen, da dieser Bereich sich danach frei bewegen könnte. Dazu müssen auch die Kupferatome in die MD-Simulation aufgenommen werden. Diese wurden in den vorliegenden Simulationen nicht miteinbezogen.

Durch die Fixierung kann keine Veränderung im aktiven Zentrum der Untereinheiten analysiert werden. Dazu muss auf zukünftige Simulationen gewartet werden, in denen die Bereiche flexibel genug sind. Eine schnelle Lösung des Fixierens bestimmter Bereiche könnte durch Anpassen des in Abb. C.34 (Seite 100) gezeigten *Chimera*-Skriptes erzielt werden. Die Datei könnte zum Fixieren der ausgewählten Atome einfach mit *Chimera* erstellt werden. Dieses Thema wird in Kapitel D.4.4.2 speziell aufgegriffen und versuchsweise gelöst (Abb. D.12 Seite 127).

Ein weiterer interessanter Versuch wäre, die Untereinheiten IIIA durch eine Untereinheit des *Panulirus interruptus*-Hämocyanins (PDBid: 1HCY) auszutauschen. Dieses bildet keine oligo-Hexamere, was in der MD-Simulation zu erkennen sein sollte. Das Sequenzalignment ergab bei *P. interruptus* im Bereich der Binde-

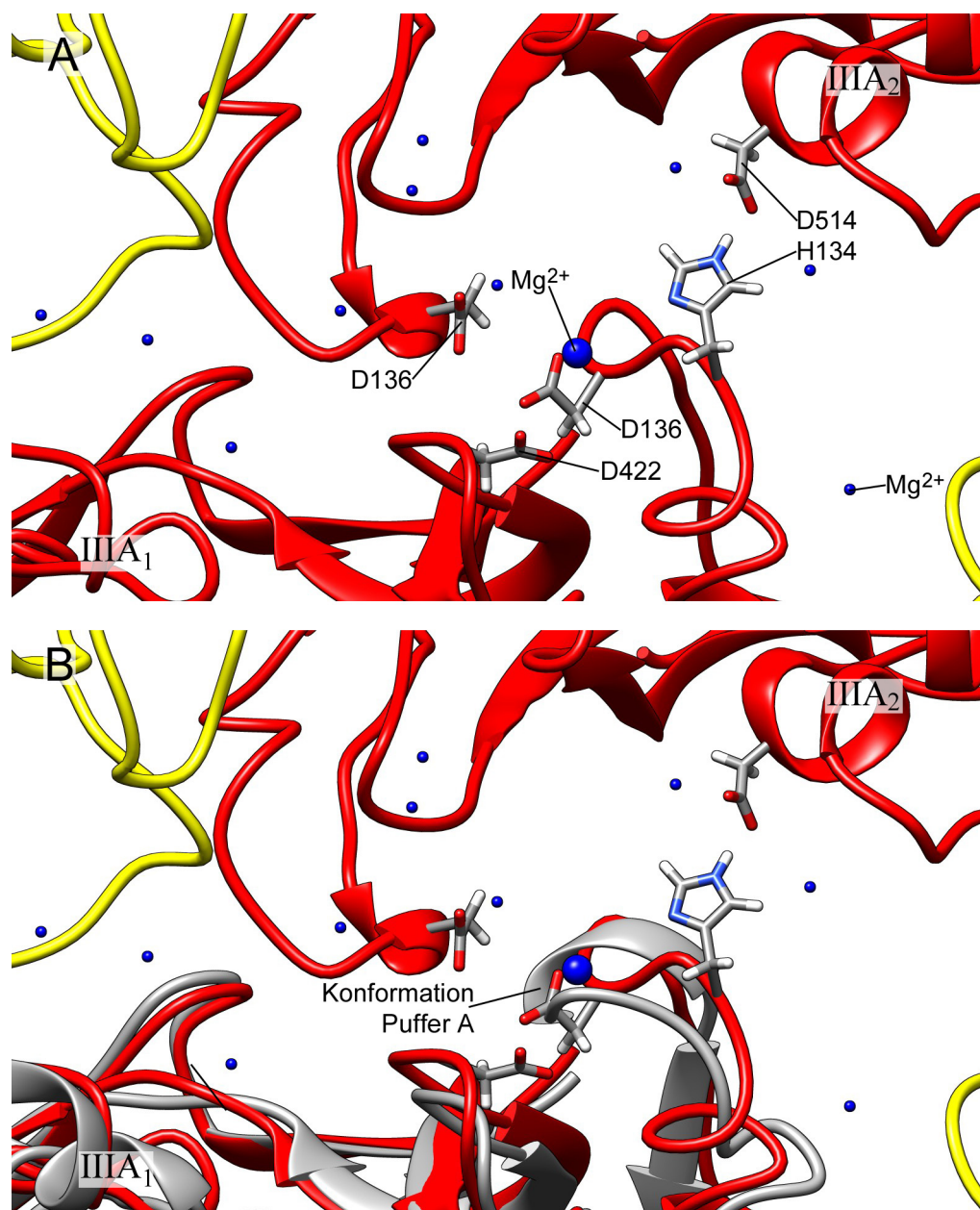


Abb. D.4: Verschiebung in der IIIA–III A-Bindestelle unter Einfluss von MgCl₂. Das Mg²⁺-Ion im Zentrum wurde zur besseren Sichtbarkeit vergrößert dargestellt. (A) Zeigt die Verbindung IIIA₁ H134 ↔ IIIA₂ D514, sowie die Aminosäuren D136 und D422 auf IIIA₁ und IIIA₂ D136. (B) Zeigt überlagert die Struktur aus Puffer A ohne MgCl₂ (grau). Hier wird durch die Verbindung IIIA₁ H134 ↔ IIIA₂ D514 die Veränderung in der Bindestelle deutlich. Der Bereich mit der Aminosäure IIIA₁ H134 wird in Richtung IIIA₂ gezogen und es entsteht Raum für das Mg²⁺-Ion.

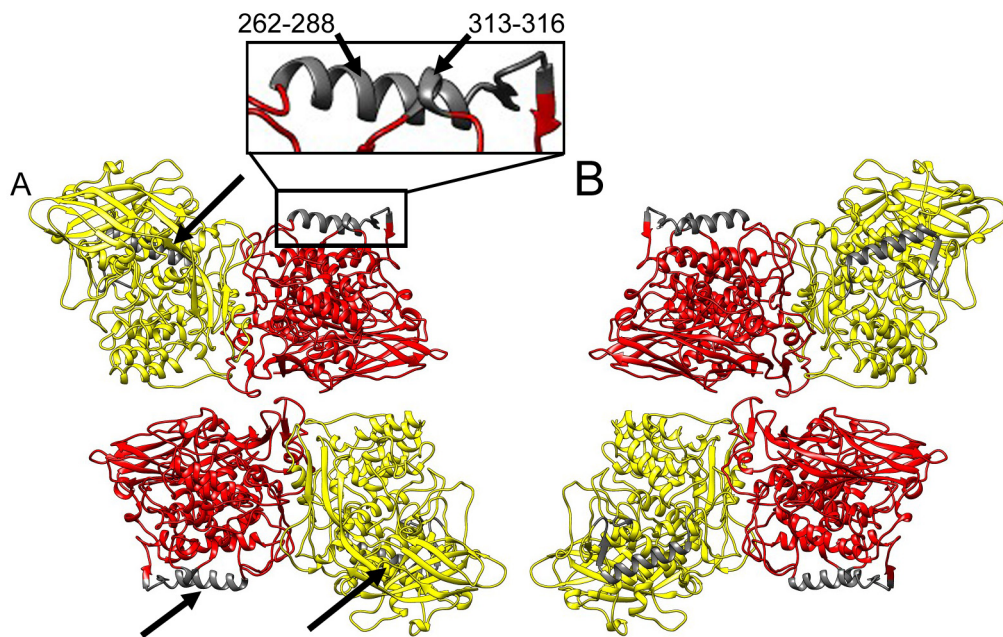
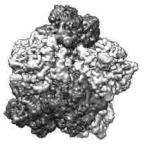


Abb. D.5: Fixierung von Bereichen für zukünftige MD-Simulationen. Es wurden exemplarisch zwei α -Helices pro Untereinheit zum Fixieren ausgewählt und hier anthrazit hervorgehoben (Pfeile). (A) Das IIIA-III A-Interface von vorne, die Pfeile zeigen auf die eingefärbten Aminosäuren 262 bis 288 sowie 313 bis 316 auf der jeweiligen Untereinheit. (B) Das IIIA-III A-Interface zur Verdeutlichung der Positionen um 180° gedreht.

Tab. D.3: Vergleich der Sequenzen der Hämocyanine von *P. interruptus* und *L. polyphemus* am IIIA–IIIA-Interface.

<i>Panulirus interruptus</i>	Anzahl saurer und basische Aminosäuren		<i>Limulus polyphemus</i>
	<i>P. interruptus</i>	<i>L. polyphemus</i>	
¹⁶² KPGTF ¹⁶⁶	1	4	¹³⁴ HPDKD ¹³⁸
⁴²⁸ E	1	1	⁴¹⁰ E
⁴⁴³ I ⁴⁵⁰ EDVEINA ⁴⁵⁰	1	3	⁴²² DHSVKVKY ⁴²⁹
⁵⁴¹ E	1	1	⁵¹⁴ D

stelle eine viel geringere Zahl geladener Aminosäuren (Tabelle D.3). Theoretisch könnte sich die bei *L. polyphemus* identifizierte Brücke H134 ↔ D514 ebenfalls bei *P. interruptus* durch die Aminosäuren K162 ↔ E541 bilden. Für die restlichen Aminosäuren scheint aber kein passender Partner in Reichweite zu liegen. Dies müsste durch eine Simulation verifiziert werden.

3.3 Vergleich der 3D-Rekonstruktionen des *L. polyphemus*-Hämocyanins

Durch den Vergleich der 7,1 Å-, 7,4 Å- und 10 Å-Rekonstruktion (siehe Abb. C.15 bis C.17) des *L. polyphemus*-Hämocyanins in dieser Arbeit konnte nicht bestimmt werden, welcher Wert für den *inner radius* ($-ir=5$ bei der 7,1 Å-Rekonstruktion und $-ir=2$ bei der 7,4 Å- und 10 Å-Rekonstruktion) die bessere Wahl darstellt. Die simulierte Dichtekarte (siehe Abb. C.28 bzw. C.29) aus den MD-Simulationen entsprach am ehesten der unsymmetrisierten 10 Å-Rekonstruktion (siehe Abb. C.17). In diesem Fall konnte man das Molekül im Raum so anordnen, dass der $-ir$ -Parameter keinen großen Einfluss auf die Rekonstruktion hatte. Zum Erreichen einer möglichst hohen Auflösung spielt die Verwendung der Symmetrisierung bei Datensätzen in dieser Größe eine entscheidende Rolle, da in den Einzelpartikeln auch immer leichte Unterschiede bestehen. Es kann zu fehlenden Untereinheiten kommen, die durch die Symmetrisierung herausgerechnet werden (Abb. D.6). Alternativ müssten deutlich mehr Einzelpartikel für die Rekonstruktion benutzt werden, um diese Defekte herauszurechnen. Somit sind die Artefakte im Bereich der Symmetrieachse definiert und müssen bei der Interpretation der Ergebnisse beachtet werden.

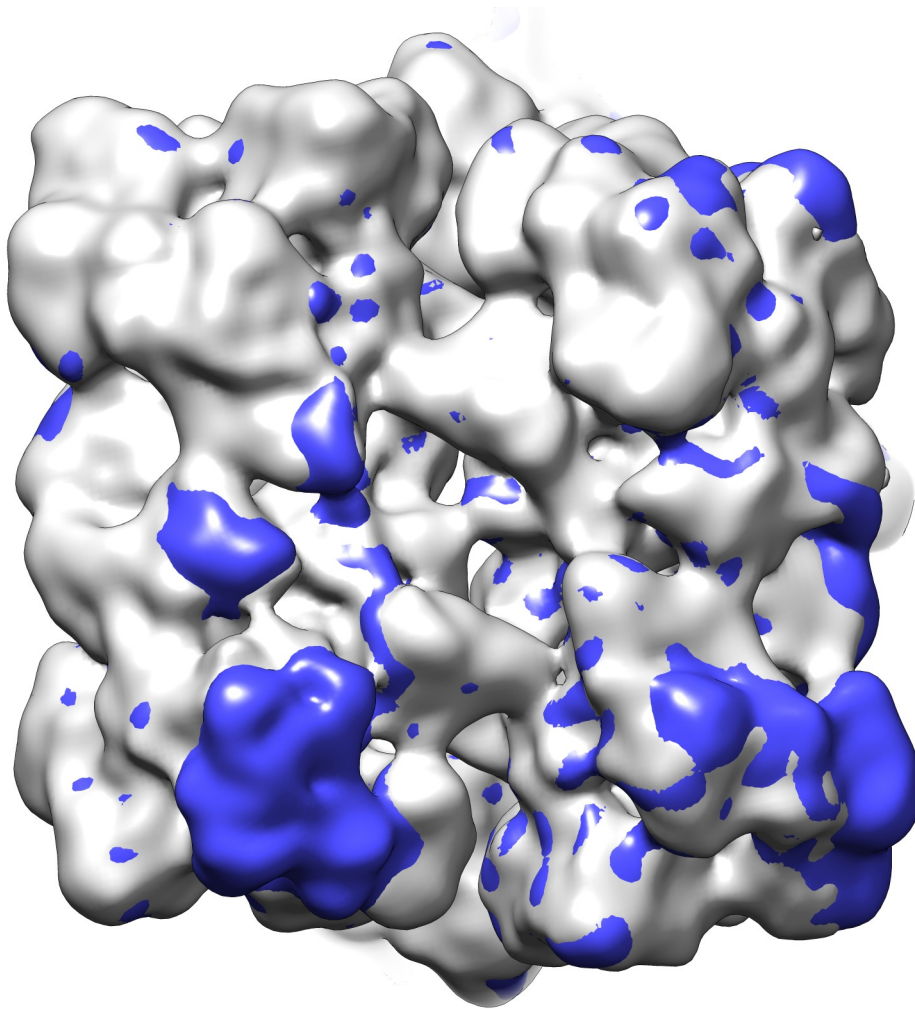
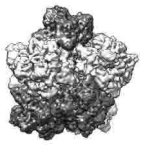


Abb. D.6: 3D-Rekonstruktion des *L. polyphemus*-Hämocyanins ohne Symmetrie (grau). Die 7,1 Å-Rekonstruktion wurde auf 10 Å gefiltert (blau) und eingepasst. Somit waren fehlende Untereinheiten (blau) zu erkennen, die auf Beschädigung von einzelnen Partikeln zurückzuführen sind.

Tab. D.4: Abmessungen der 7,1 Å-Rekonstruktion und der 12 Å-Rekonstruktion aus Martin et al. (2007).

	12 Å-Rekonstruktion	7,1 Å-Rekonstruktion
<i>x</i> -Achse	284 Å	260 Å
<i>y</i> -Achse	271 Å	267 Å
<i>z</i> -Achse	282 Å	254 Å

Tab. D.5: Abmessungen des *L. polyphemus*-Hämocyanins, übernommen aus Martin et al. (2007).

1x6mer im 2x6mer		2x6mers im 4x6mer			4x6mers im 8x6mer		Referenz
Distanz (Å)	Rotation (°)	Distanz (Å)	Dihexamer-verschiebung (Å)	Kippwinkel (°)	Distanz (Å)	Staffelungswinkel (°)	
100 ± 3	105 ± 5	102 ± 3	14 ± 1	16 ± 2.5	106 ± 3	42 ± 2	Martin et al. (2007)
99 ± 2	105 ± 2	93 ± 2	12 ± 1	14 ± 3	104 ± 2	40 ± 2	diese Arbeit

3.4 Abmessungen des Hämocyanins

Durch die Überlagerung der 7,1 Å-Rekonstruktion mit der 12 Å-Rekonstruktion von Martin et al. (2007) (EMDataBank¹ Kennnummer 1304) konnten die beiden Dichtekarten direkt verglichen werden (Abb. D.7). Hierbei fiel auf, dass die 7,1 Å-Rekonstruktion deutlich kompakter war als die 12 Å-Rekonstruktion (Tabelle D.4). Das wirkt sich auch auf das pseudoatomare Modell aus, in dem die acht Hexamere jetzt mehr zusammen gerückt sind. Somit konnten die unterschiedlichen Werte für die Distanz der 2x6mere innerhalb eines 4x6mers erklärt werden (Tabelle D.5). Die Verkleinerung des Kippwinkels war bedingt durch die unsymmetrischen Änderungen der Abmessungen des Hämocyanins (siehe Tabelle D.4). Beim starren Einpassen der Hämocyaninuntereinheiten rückten diese zwar näher zusammen (Abb. D.8 und D.9), eine Berührung der Seitenketten (*clashes*) des pseudoatomaren Modells wurde jedoch durch das flexible *Fitting* mit *Direx* vermieden.

¹Eine Internetquelle für 3D-EM-Rekonstruktionen (Lawson et al., 2011)

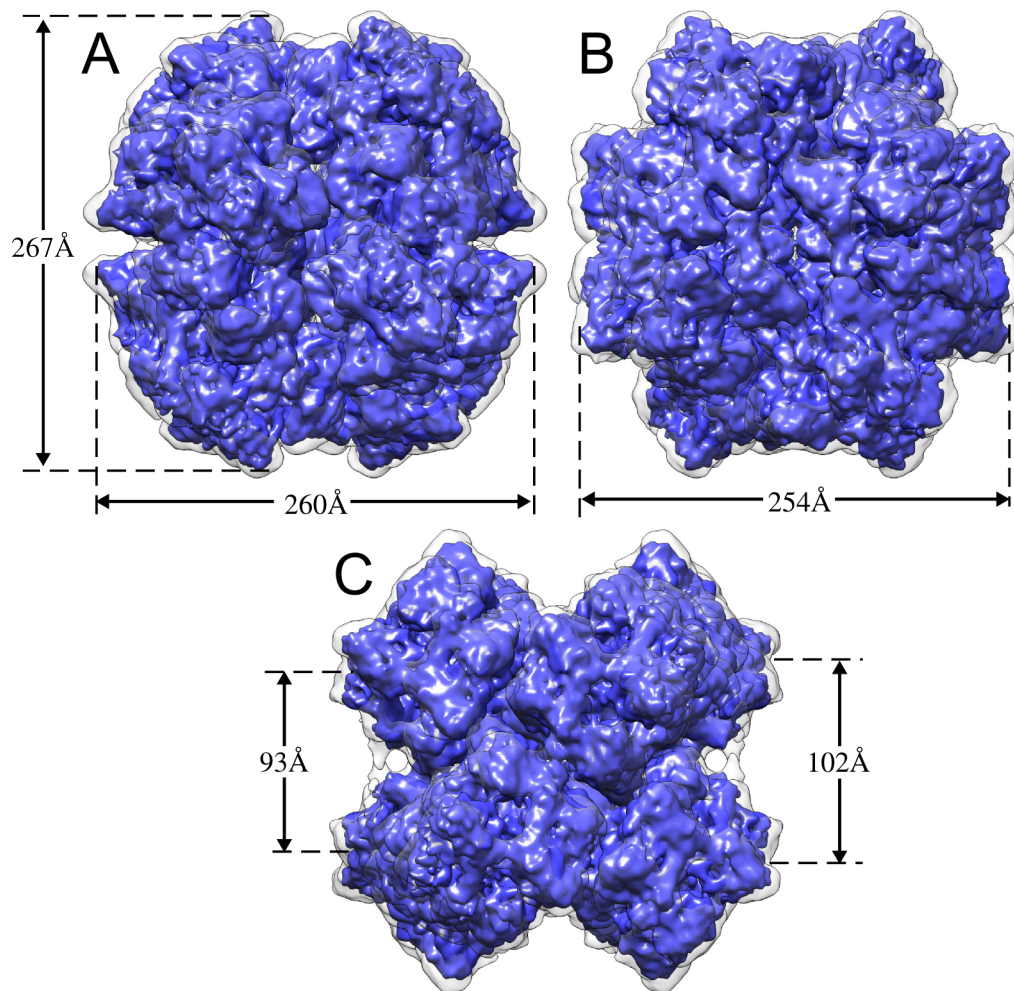
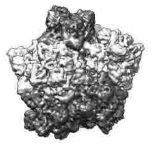


Abb. D.7: Vergleich der 3D-Rekonstruktionen des *L. polyphemus*-Hämocyanins zwischen Martin et al. (2007) (transparent) und dieser Arbeit (blau) bei gleichem Massenschwellenwert. (A) Blick entlang der z -Achse mit den Abmessungen 260 \AA (x -Achse) und 267 \AA (y -Achse). (B) Drehung um die y -Achse um 90° somit Abmessung entlang der z -Achse mit 254 \AA . (C) Blick entlang der y -Achse. Durch die kompakteren Abmessungen (siehe Tabelle D.4) änderte sich die Distanz der 2x6mere von 102 \AA (Martin et al., 2007) auf 93 \AA ($7,1 \text{ \AA}$ -Rekonstruktion).

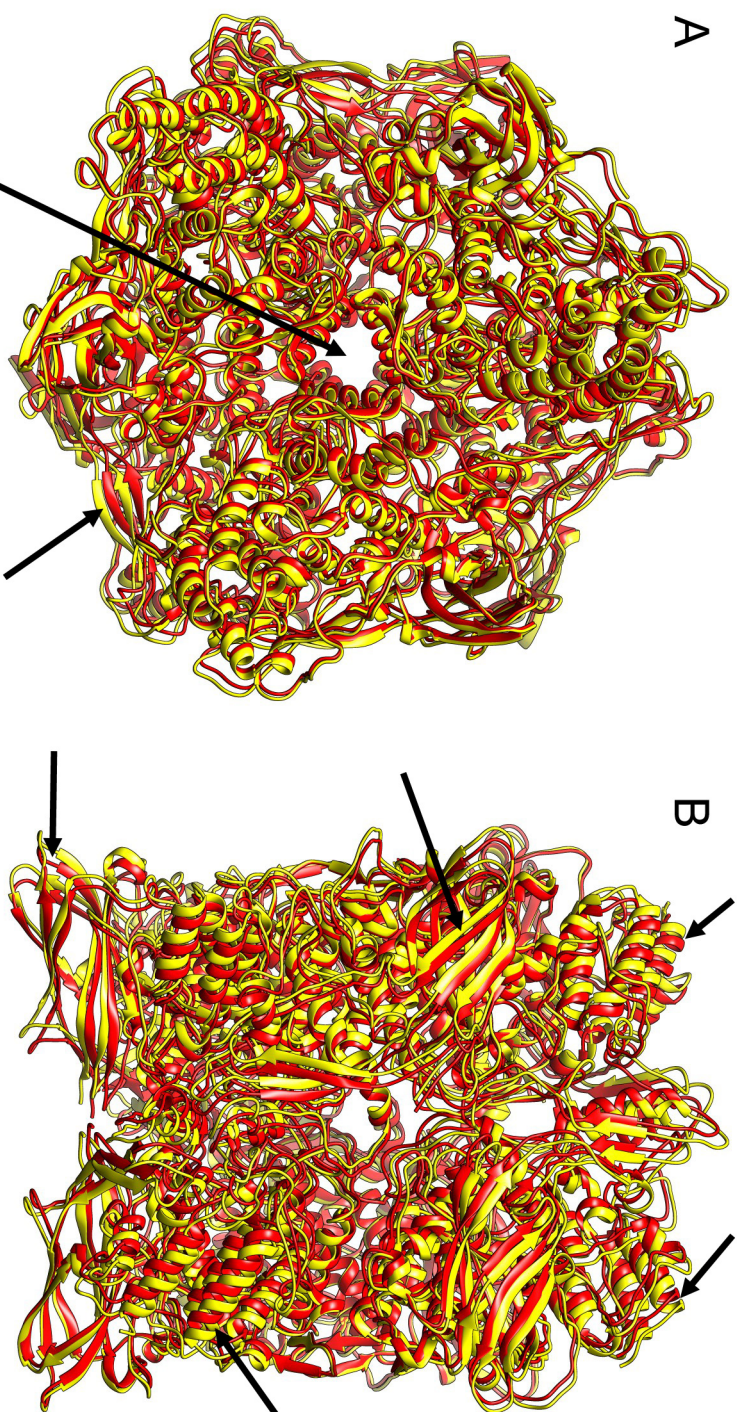


Abb. D.8: Vergleich eines extrahierten Hexamers des pseudoatomaren Modells der 7,1 Å-Rekonstruktion des *L. polyphemus*-Hämocyanins (rot) mit einem Homologiemodell, das mithilfe des kristallographischen Hexamers von *P. interruptus*-Hämocyanin (PDBid 1HCY) erstellt wurde (gelb). Dabei wurden die Untereinheiten von *P. interruptus* mit dem Chimera-Modul MatchMaker durch die Untereinheiten von *L. polyphemus* ersetzt. (A) Aufsicht, (B) Seitenansicht. Wie man sieht, ist die rote Struktur etwas kompakter (Pfeile), was darauf hinweist, dass die 7,1 Å-Rekonstruktion minimal zu klein sein könnte. Die Abweichung ist aber geringer als das Auflösungsvermögen der 3D-Rekonstruktion und daher unproblematisch. Abgesehen davon überlagern sich die Untereinheiten perfekt, was vermuten lässt, dass die Assoziation zum 8x6-mer keine Deformation der Hexamere bewirkt.

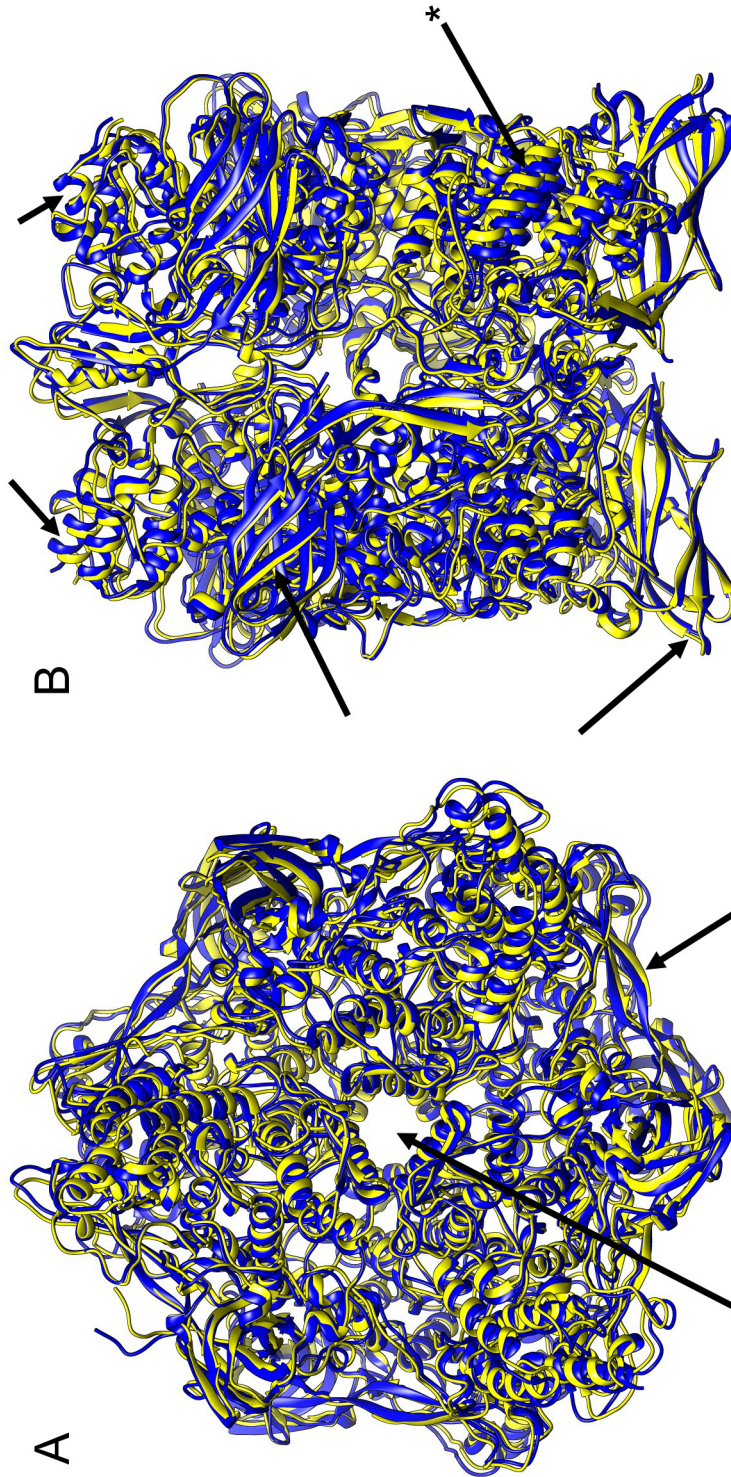
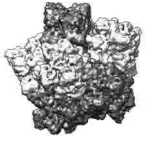


Abb. D.9: Vergleich eines extrahierten Hexamers des pseudoatomaren Modells der 12 Å-Rekonstruktion des *L. polyphemus*-Hämocyanins aus Martin et al. (2007) (blau) mit einem Homologiemodell, das mithilfe des kristallographischen Hexamers von *P. interruptus*-Hämocyanin (PDBid 1HCY) erstellt wurde (gelb). Beide Strukturen besitzen in etwa die gleiche Größe. Lediglich die mit * markierte Untereinheit verschiebt sich etwas zum Zentrum hin. (A) Aufsicht, (B) Seitenansicht.

4 Programmierarbeiten

4.1 *Graphical user interface (GUI)*

Die GUI wurde in Python programmiert, für die grafische Darstellung kam das Paket PyQt (Riverbank-Computing, 2013) zum Einsatz. Sowohl die Programmiersprache als auch das Grafikpaket PyQt waren betriebssystemübergreifend verwendbar (Summerfield, 2007). Allerdings war die Darstellung in den Betriebssystemen unterschiedlich. Hier konnte unter Umständen die Menüstruktur im oberen Bildschirmrand dargestellt werden und nicht, wie in dieser Arbeit demonstriert, im Programmfenster. Dieses Phänomen war in diesem Kontext nicht relevant, da die GUI immer auf dem Rechen-Cluster der Arbeitsgruppe ausgeführt wurde.

Die GUI befand sich derzeit in einem frühen Beta-Stadium. Das Ziel der einfachen Erweiterbarkeit wurde mithilfe von Konfigurationsdateien erfüllt. Dabei wurde jeweils ein Werkzeug durch eine Konfigurationsdatei in die GUI eingebunden. Die Kategorisierung und Darstellung in der GUI funktionierte danach automatisch.

Durch Beta-Tests waren bereits erste Fehler identifiziert worden. Das Ändern des Arbeitsverzeichnisses funktionierte, allerdings wechselte die Anzeige wieder zum ursprünglichen Arbeitsverzeichnis zurück. Man konnte nur durch die Ausgabe in der Kommandozeile verfolgen, in welchem Verzeichnis man gerade arbeitete.

Ein Nachteil der GUI war, dass sie keine Art von Fehlerbehandlung besaß. Darauf wurde zu Gunsten der einfachen Erweiterbarkeit verzichtet. Der Benutzer war somit selbst dafür verantwortlich zu überprüfen, ob die gemachten Eingaben dem erwarteten Format entsprachen. Die GUI gab lediglich Hinweise über das erwartete Format.

Für die Zukunft sind hier mehrere Erweiterungen geplant, die das Arbeiten mit der GUI weiter erleichtern sollten. Derzeit wurden alle Meldungen der ausgeführten Software-Tools in der Kommandozeile ausgegeben. Hier sollte innerhalb der GUI ein Textfeld erscheinen, in dem die Informationen dargestellt werden, um die Übersichtlichkeit zu steigern. Es wurde auch über einen Befehlsverlauf nachgedacht, der die ausgeführten Kommandos mit Parametern und Datum anzeigen kann.



4.2 Administration am Rechen-Cluster

Die Umstellung auf ein Mehrbenutzersystem bringt Vor- und Nachteile mit sich (Frisch, 2003). Dank des *Queueing*-Systems (Cluster Resources, 2009) konnten die verfügbaren Ressourcen gleichmäßig auf die verschiedenen Benutzer verteilt werden. Die Hardware konnte dadurch effizienter arbeiten.

Der administrative Mehraufwand sollte nicht unterschätzt werden. Das zur Administration des Rechen-Clusters nötige *Know-How* wurde in mehreren DokuWiki-Artikeln zusammengefasst und ständig erweitert. Der Wiki kann jedoch die Erfahrung eines Administrators nicht ersetzen, da viele auftretende Probleme (noch) nicht durch die DokuWiki-Artikel abgedeckt wurden.

4.3 DokuWiki

Die Software DokuWiki wurde von der Seite www.wikimatrix.org (Hüttemann and Riebesell, 2013) hinter MediaWiki² derzeit auf Platz 2 der meist verglichenen Wiki-Seiten gelistet. DokuWiki wurde hier gewählt, da es besonders einfach zu benutzen ist und keine besonderen Systemvoraussetzungen benötigt. Die Software arbeitet mit reinen Textdateien zur Darstellung des Inhalts und verzichtet dabei auf komplexe Datenbanksysteme (Gohr, 2013).

Eine Wiki-Software eignet sich besonders gut als Wissensdatenbank in kleinen Arbeitsgruppen und Firmen (Figura and Gross, 2013). Voraussetzung ist allerdings die Motivation zur Mitarbeit an den Wiki-Artikeln. Dies kann durch besonders gut strukturierte Artikel erreicht werden, in denen die Information klar gegliedert ist und von den Benutzern schnell gefunden wird.

Die DokuWiki-Software wurde in der Standardversion installiert, ohne besondere Zusatzmodule. Mittlerweile gibt es spezielle Erweiterungen, die den Benutzer darauf hinweisen, wenn die Qualität des Beitrags minderwertig ist (Gohr et al., 2010). Hier wird anhand verschiedener Statistiken versucht (z. B. Verhältnis Wörter zu Überschriften, Anzahl von Überschriften, etc.) dem Benutzer eine direkte Rückmeldung und Tipps zu geben, ob der gerade erstellte Artikel qualitativ hochwertig ist. Der Benutzer kann dann seinen Beitrag anpassen.

Diese Art von Modulen kommt in unserem DokuWiki nicht zum Einsatz. Nach diesen Maßen bewertet, würde die Implementierung in unserer Arbeitsgruppe

²MediaWiki ist die frei zugängliche Wiki-Software, bekannt aus dem Wikipedia Projekt.

nicht so gut abschneiden, da teilweise nur stichpunktartig Details dokumentiert wurden und kein zusammenhängender Fließtext entstand. Da es sich aber um recht überschaubares Wissen handelt, konnte unser Wiki seine Aufgabe recht gut erfüllen. Er wurde dank geschützter und privater Bereiche zusätzlich als privates Laborbuch benutzt.

4.4 Entwickelte Software-Tools

Allgemein ist zu den entwickelten Software-Tools zu sagen, dass ein ausführlicher Test nur innerhalb unserer Arbeitsgruppe stattgefunden hat. Dies reicht nicht aus, um alle Software-Fehler aufzudecken, da sich feste Routinen eingespielt haben und die Tools alle nach einem bekannten Muster ausgeführt wurden. Dazu wäre ein Test von anderen Arbeitsgruppen äußerst hilfreich.

Die Fehlerbehandlung war bei den meisten Programmen nicht ausreichend implementiert. Vom Benutzer wurde erwartet, sich an die Anleitung im entsprechenden DokuWiki-Artikel zu halten, bei Missachtung waren Fehler zu erwarten. Der Benutzer war somit angehalten, die Ergebnisse der Programme zu überprüfen.

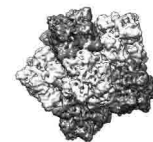
4.4.1 Allgemeine Software-Tools

Die Ausgabe von `readTifHead.py` (Abb. D.10) war für die Darstellung einer großen Anzahl von Bildern sehr unpraktisch. In Zukunft soll eine tabellarische Darstellung folgen, mit einem Datensatz pro Zeile.

4.4.2 Software-Tools zum Arbeiten mit MD-Simulationen

Für die Erstellung der Graphen zur MD-Simulations-Analyse wurden die Programme `gromacsgraph.py`, `gromacshist.py` und `gromacsdisc.py` entwickelt. Generell können solche Diagramme auch über eine Tabellenkalkulationssoftware wie `Microsoft Excel` erstellt werden. Wenn jedoch viele Dateien bearbeitet werden sollen, würde die Kalkulation in `Excel` zu aufwändig werden, da die Ergebnisse der Messung mit dem `Gromacs`-Modul `g_dist` keine direkt in `Excel` verwertbare Ausgabe produzieren.

Durch die Verwendung der genannten Programme konnte die Zeit für nahezu beliebig viele Messungen auf wenige Sekunden reduziert werden. An dieser Stelle



```
1 Imagename: DpHb_1zu5_w_002.tif
      high tension: 120000.0V
3           TEM mode: 0 => bright field
      Image distance X: 0.436119645834nm
5           Image distance Y: 0.436119645834nm
      Pixel sizes: (15600.0, 15600.0)um
7 current magnification relative to plate: 49000.0x
      corrected magnification: 1.0x
9           post magnification: 1.46000003815
      Image size X: 2048px
11          Image size Y: 2048px
      Image Name: DpHb_1zu5_w_002
13          Image Folder: Images
      Camera type: F416
15          Type of TEM: MCL_02
      Microscope information:
17          Specimen information:
      TEM aberrations: (0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0)
19          Stage position, X Y Z A B: (0.00025566399563103914, 2.2401000023819506e-05,
4.6665201807627454e-05, -0.00048800636432133615, 0.0)
      Creation date: 131859734
21          Creation time: 76723
      Exposure time: 1000.0ms
23          Dose: 645.048706055e/A^2
      data type: 1 => unsigned char (8 bit)
```

Abb. D.10: Die Ausgabe von `readTifHead.py` ist zur Zeit sehr unübersichtlich. Hier ist für die Zukunft eine tabellarische Ausgabe mit den wichtigsten Werten geplant, damit die Informationen für eine Vielzahl von Bildern übersichtlich dargestellt werden kann.

war eine geschickte Wahl des Dateinamens besonders von Vorteil, da die genannten Programme *wildcards*³ erkennen. Somit konnte das Diagramm für eine Gruppe von Messdateien mit einem einzigen Befehl erstellt werden.

Zur Zeit ist die Größe der erstellten Grafiken auf DIN-A4 Seiten optimiert. Dadurch ergeben sich bei der Verwendung von `gromacsgraph.py` Darstellungsprobleme bei sehr langen MD-Simulationen mit vielen Messwerten (>10 000). Die Grafiken werden mit einer Auflösung von 300 DPI und einer Breite von 30 cm erstellt. Es ergeben sich etwa 3 600 Bildpunkte (ca. 12 in × 300) zur Darstellung des Graphen. Die Anzahl der Messpunkte wurde derzeit um einen festen Faktor reduziert, indem immer über 10 Messwerte gemittelt wurde. Hierdurch

³Eine *wildcard* ist ein Platzhalter und steht für eine beliebige Zeichenfolge im Dateinamen. Sie wird durch ein „*“ repräsentiert. Damit können automatisch alle Dateien ausgewählt werden, die mit gleicher Buchstabenfolge beginnen (z. B. `intra-hex*`) oder enden (z. B. `*.xvg`).

ging leider die Dynamik verloren (Abb. D.11), die in Zukunft durch Balken mit Standardabweichung ersetzt werden soll.

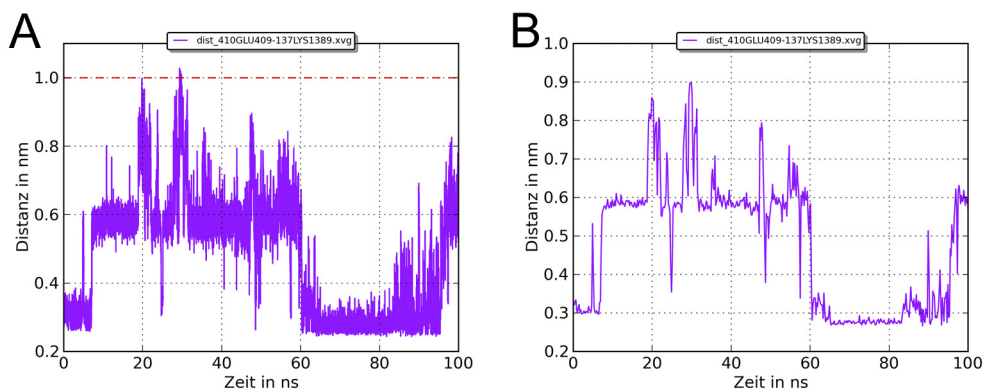
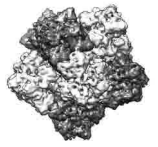


Abb. D.11: Lesbarkeit der Graphen von `gromacsgraph.py`. (A) Graph mit 25 000 Messpunkten. (B) Graph mit 500 Messpunkten, gemittelt über 50 Messpunkte des selben Datensatzes wie in (A). Es wird deutlich, wenn in (A) die Messwerte um 3 \AA pro Zeiteinheit variieren, werden in (B) die Mittelwerte angezeigt und es die Information der Dynamik geht verloren. Die Grafik sollte in Zukunft mit Fehlerbalken erweitert werden, um diese Information zu erhalten, den Graphen aber dennoch übersichtlich zu machen.

Bei der Erstellung des Balkendiagramms mit dem Programm `gromacshist.py` wurde für die Simulationszeit eine feste Höhe von 55 % der y -Achse gewählt, die eigentlichen Prozent-Werte stehen zusätzlich in den Balken. Aufgrund der gleichen Höhe der y -Achse waren alle Diagramme miteinander vergleichbar, falls sich jedoch eine Entfernungsmessung über 55 % der Simulationszeit in einer bestimmten Distanz befand, wurden die Extremwerte beschnitten oder fielen optisch nicht direkt auf. Um viele Messungen übersichtlich auf engem Raum darzustellen waren hierfür Scheibendiagramme mithilfe von `gromacsdisc.py` eine kompakte Lösung. Microsoft Excel bietet keine vergleichbare Möglichkeit. Die Programme `gromacshist.py` und `gromacsdisc.py` arbeiteten ebenfalls mit *wildcards* und konnten in Sekunden nahezu beliebig viele Dateien auswerten.

Das Skript `restrain35angst.sh` ist kein vollständig ausgearbeitetes Programm, das auf beliebige PDB-Dateien angewendet werden kann, um bestimmte Atome für die MD-Simulation zu fixieren. Vor seinem Einsatz für andere Projekte müssten Dateinamen, Radius sowie das Zentrum des Fixierradius editiert werden. Sie



```

1 fh = open("/PATH/TO/FOLDER/restrain.itp","w") # Restrain Datei oeffnen
2 fh.write("[ position_restraints ]\n") # Header schreiben
3 for j in chimera.selection.currentAtoms(): # fuer jedes ausgewaehlte Atom
4     fh.write("%s 1 1000 1000 1000\n" % (j.serialNumber)) # schreibe Atom in die
5     # Datei, gefolgt von den Parametern fuer Gromacs
6 fh.close() # Datei wieder schliessen

```

Abb. D.12: Fixieren ausgewählter Atome mithilfe von Chimera. Hierzu musste in der ersten Zeile der Text ‘PATH/TO/FOLDER‘ durch den entsprechenden Pfad zum Arbeitsverzeichnis angepasst werden.

sind fest im Skript eingetragen. In Zukunft soll mit Übergabeparametern gearbeitet werden. Somit könnten Radius, drei Raumkoordinaten zur Definition des Zentrums sowie die Dateinamen, auf die das Skript angewendet wird, direkt übergeben werden. Erst nach dieser Änderung würde das Programm auch in die GUI aufgenommen werden und könnte universell für neue Projekte eingesetzt werden. Es bleibt an dieser Stelle aber weiterhin offen, ob das Fixieren mit einem festen Radius um einen Punkt Sinn macht. Zukünftige MD-Simulationen werden hier hoffentlich Antworten liefern. Alternativ zum Fixieren mit `restrain35angst.sh` könnte man auch ein Programm für Chimera schreiben, in dem man die zu fixierenden Aminosäuren mit der Maus auswählt und diese sodann in eine Datei geschrieben würden. Das ist für die Zukunft geplant und wurde zu Testzwecken in Abb. D.12 versucht. Ausführliche Tests müssten hier noch durchgeführt werden.

Das Programm `md2volume.sh` arbeitete nur auf einer Atom-Trajektorien-Datei. Zur Erstellung der 3D-Volumina aus einer Serie von MD-Simulationen musste das Programm für jede Simulation ausgeführt werden. Die Ergebnisse mussten danach mit dem EMAN2-Programm `e2proc3d.py` kombiniert werden. Die Weiterverarbeitung wie das Filtern der 3D-Volumina geschah ebenfalls mit verschiedenen EMAN2-Modulen.

Um mit Gromacs die einzelnen Atome zu vermessen, wurde eine Index-Datei benötigt. Sie enthält Namen von Atomen bzw. Atomgruppen, gefolgt von den Atomnummern. Das vorgestellte Chimera-Skript (Abb. C.34 Seite 100) erstellte die Namen basierend auf der ausgewählten Aminosäure mit zugehöriger Kette und den entsprechenden Atomnummern automatisch nach Auswahl der entsprechenden Atome in Chimera. Dieser Prozess ersparte das manuelle Erstellen der

Atomgruppen und verhinderte eventuelle Übertragungsfehler. Hier wäre es Dank eines ausgereiften *plugin*-Systems von **Chimera** möglich, ein Modul zur Integration in **Chimera** zu erstellen (Pettersen et al., 2004; Goddard et al., 2005; Heiland et al., 2007). Dies ist für die Zukunft geplant, um den Weg über die manuelle Eingabe in IDLE zu umgehen.

4.4.3 Software-Tools zum flexiblen Fitten

Das Programm `hbond2direx` besitzt keine Fehlerkontrolle. An dieser Stelle ist es wichtig, dass man sich beim Erstellen der `hbond`-Datei mithilfe von **Chimera** genau an die Anleitung im DokuWiki hält. Andernfalls ist die Datei eventuell nicht korrekt konvertierbar. Das Programm gibt in diesem Fall keine Fehlermeldung aus. Der Fehler würde erst bei der Verarbeitung mit **Direx** auffallen. Hier soll in Zukunft eine Prüfung eingebaut werden, die den Benutzer auf einen Fehler hinweist.

4.4.4 Software-Tools zum Erstellen und Bearbeiten von 3D-Rekonstruktionen

`plotfsc.py` erstellt eine Grafik, die nicht genau einer mathematischen Norm entspricht. Die Werte der Auflösungskriterien werden automatisch in die Grafik eingetragen, allerdings nicht an die erwartete Position (Abb. D.13). Es ist geplant, die Grafik entsprechend Abb. C.11 (Seite 66) anzugleichen, damit sie ohne manuelle Anpassungen in Publikationen verwendet werden kann.

`anglespread2.py` ist an das EMAN2-Modul `e2eulerexplor.py` (Tang et al., 2012) angelehnt, das eine vergleichbare 3D-Karte erzeugt. Im Gegensatz zu `anglespread2.py` bietet das EMAN2-Modul mehrere Zusatzfunktionen. So zeigt es direkt die entsprechenden Klassensummen aus der jeweiligen Projektionsrichtung an. Allerdings arbeitet es nur mit Datensätzen, die mit EMAN2 prozessiert wurden. `anglespread2.py` verzichtet auf die Anzeige der Klassensummen, arbeitet aber mit Datensätzen aus den 3D-Rekonstruktionsprogrammen **IMAGIC5** (van Heel et al., 1996), **EMAN** (Ludtke et al., 1999) und **Sparx** (Hohn et al., 2007).

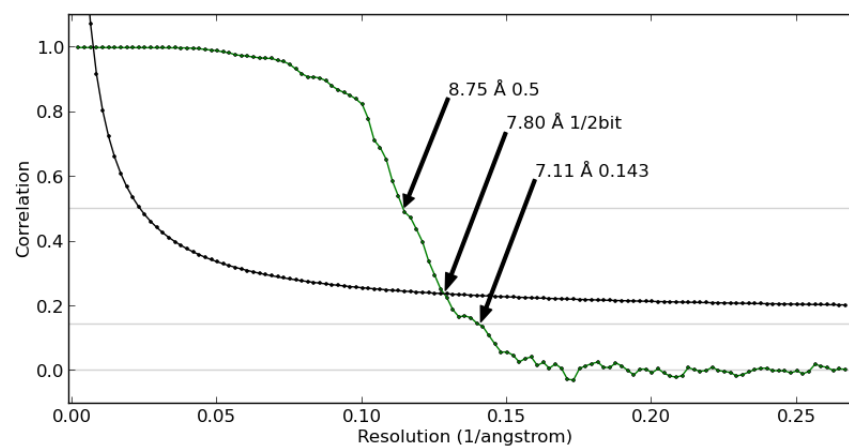


Abb. D.13: FSC erstellt mit `plotfsc.py`. Die Auflösungen werden direkt in der Grafik angezeigt, jedoch zeigt die Pfeilspitze auf den Punkt an dem sich die Kurven schneiden und nicht auf die x -Achse, an der die Auflösung abgelesen wird.



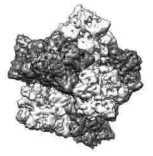
Zusammenfassung

Die Transmissionselektronenmikroskopie gepaart mit bioinformatischen Methoden zur digitalen Bildverarbeitung ist ein schneller Weg zur Erstellung dreidimensionaler Rekonstruktionen großer Proteinkomplexe. Durch die Kombination der 3D-Elektronenmikroskopie mit der Röntgenstruktur von Untereinheiten erhält man ein pseudoatomares Modell der Quartärstruktur.

In dieser Arbeit wurden auf diese Weise die Quartärstrukturen von drei unterschiedlichen respiratorischen Proteinen analysiert (einem Hämoglobin und zwei Hämocyaninen). Zudem wurden spezielle Software-Tools entwickelt, um vorhandene Softwarepakete besser miteinander kombinieren zu können.

Die ca. 15 Å 3D-Rekonstruktion des Hämoglobins vom Wasserfloh *Daphnia pulex* klärt die umstrittene Frage, wie viele Untereinheiten die Quartärstruktur aufbauen: Es sind 16 (mit je zwei Häm-Domänen), angeordnet in zwei Schichten als D4-symmetrisches *Sandwich*. Die ca. 15 Å 3D-Rekonstruktion des 2x6meren Hämocyanins des Flusskrebses *Astacus leptodactylus* gibt neue Einblicke in die Kontaktstelle zwischen den beiden Hexameren; sie liegt im Bereich der Domäne 3. Bei dem aus 48 Untereinheiten bestehenden Hämocyanin des Pfeilschwanzes *Limulus polyphemus* wurde eine Auflösung von ca. 7 Å erreicht. Die Homologiemodelle der Untereinheiten wurden flexibel *gefittet*. An einer der Kontaktstellen zwischen den beiden Halbmolekülen wurden Molekulardynamik (MD)-Simulationen durchgeführt, um mehr über die Art der chemischen Bindung an dieser Kontaktstelle zu erfahren.

Speziell für die Kombination von 3D-Elektronenmikroskopie und MD-Simulation wurden verschiedene bioinformatische Werkzeuge und eine leicht zu erweiternde universelle grafische Benutzeroberfläche (GUI) entwickelt.



E Appendix

1 2 %ige Uranylacetat Lösung

1 g Uranylacetat auf 50 ml destilliertes Wasser 5 Minuten schütteln

2 2 %ige Uranylformiat Lösung

Destilliertes Wasser kochen und auf Raumtemperatur abkühlen lassen

0,1 g Uranylformiat in ein leeres Regenzglas geben

5 ml Destilliertes Wasser hinzugeben

10 Minuten in Dunkelheit schütteln

2 µl 10 M Kaliumhydroxid pro 1 ml Lösung

10 Minuten in Dunkelheit schütteln

10 Minuten in Tischzentrifuge bei maximaler Geschwindigkeit zentrifugieren

Die Lösung durch ein 0,2 µm Filter in Dunkelheit filtern

3 Negativfilme entwickeln

Das Entwickeln der Filme ist ein zweistufiger Prozess. Zunächst werden die Filme für 15 Minuten entwickelt. Nach intensivem Spülen mit frischem Wasser werden sie für weitere 12 Minuten fixiert. Nach einer gründlichen Trocknung können die Negativfilme gescannt werden.

4 Pufferrezepte

Daphnia pulex Puffer

Puffer zur Verdünnung

50 mM TRIS Puffer mit pH 8

Puffer für Chromatofokussierung

Puffer A: 50 mM Natriumphosphatpuffer mit pH 8,1

Puffer B: 50 mM Essigsäure mit pH 4,2

Astacus leptodactylus Puffer

Puffer zur Verdünnung und Gelchromatographie

100 mM TRIS, 10 mM Magnesiumchlorid, 10 mM Calciumchlorid mit pH 7,6

Limulus polyphemus Puffer

Puffer zur Verdünnung und Gelchromatographie

100 mM TRIS, 10 mM Magnesiumchlorid, 10 mM Calciumchlorid mit pH 7,8

5 Graphen der MD-Simulation

Im Ergebnisteil wurde zugunsten der Scheibendiagramme auf die Graphen der Aminosäurenabstände verzichtet. Diese werden im Appendix aufgeführt.

Die einzelnen Graphen tragen aufgrund der automatischen Auswertung jeweils die Dateinamen als Beschriftung. Die Dateinamen sind immer nach dem gleichen Prinzip aufgebaut. Der Dateiname „`dist_136D_A-423H_C.xvg`“ steht Z. B. für:

`dist_` bedeutet Distanzmessung.

`136D` bedeutet Aminosäure 136 vom Typ Asparaginsäure (D)

`_A` ist die Kennung der Polypeptidkette in Chimera. In diesem Fall steht A für Untereinheit vom Typ IIIA₁.



-- bedeutet Verbindung zu...

423H bedeutet Aminosäure 423 vom Typ Histidin (H)

_C ist die Kennung der Polypeptidkette in **Chimera**. In diesem Fall steht C für Untereinheit vom Typ IIIA₂.

.xvg Ist die Dateiendung.

Um Platz zu sparen wurde bei der Umwandlung der Graphen zu Scheibendiagrammen der Dateiname sowohl am Anfang als auch am Ende beschnitten. Es blieb lediglich der Kern übrig: „136D_A- -423H_C“.

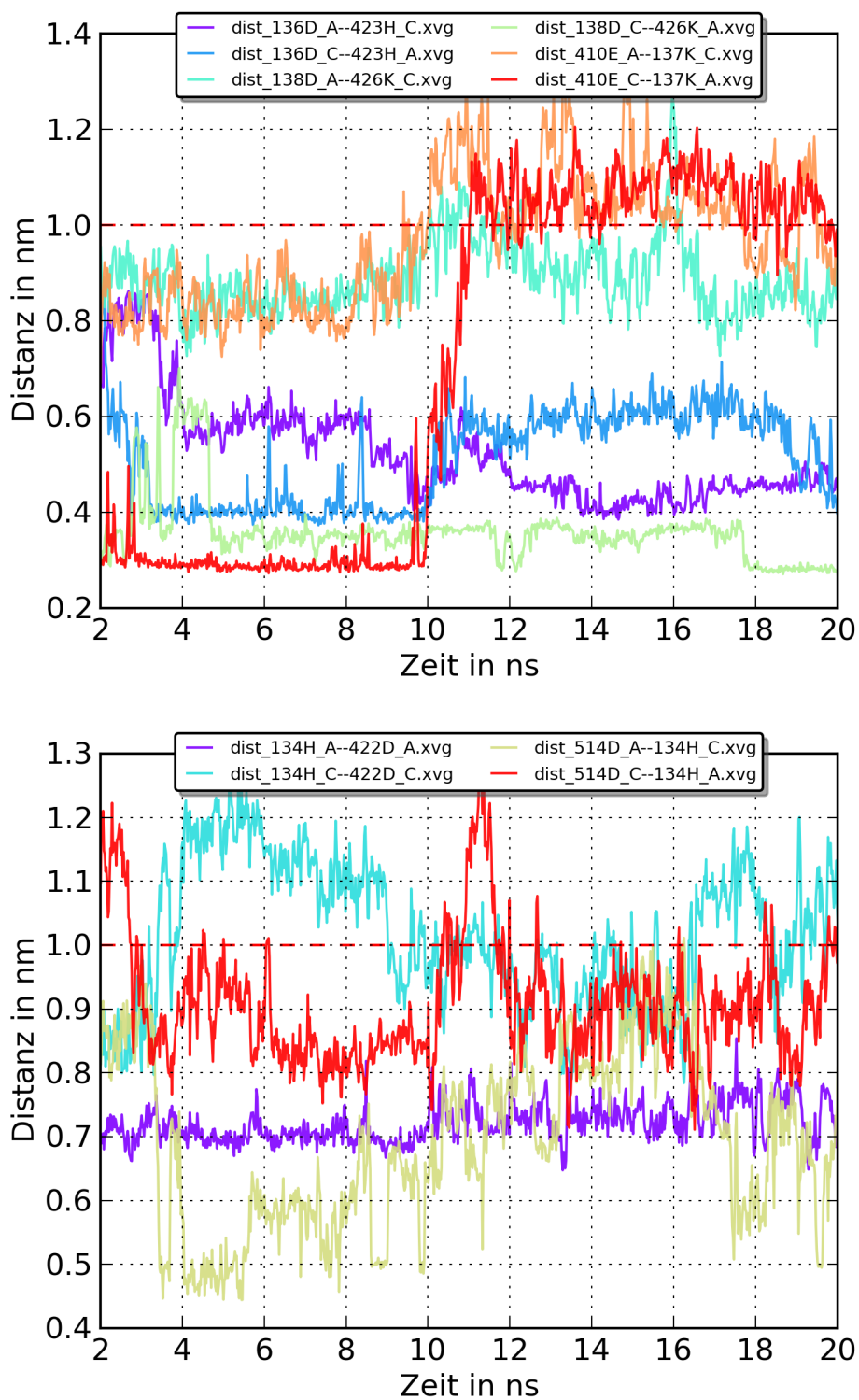


Abb. E.1: Aminosäurenabstände der 1. MD-Simulation im Puffer A mit 20 ns. Die Graphen wurden in das Scheibendiagramm in Abb. E.2 umgewandelt.

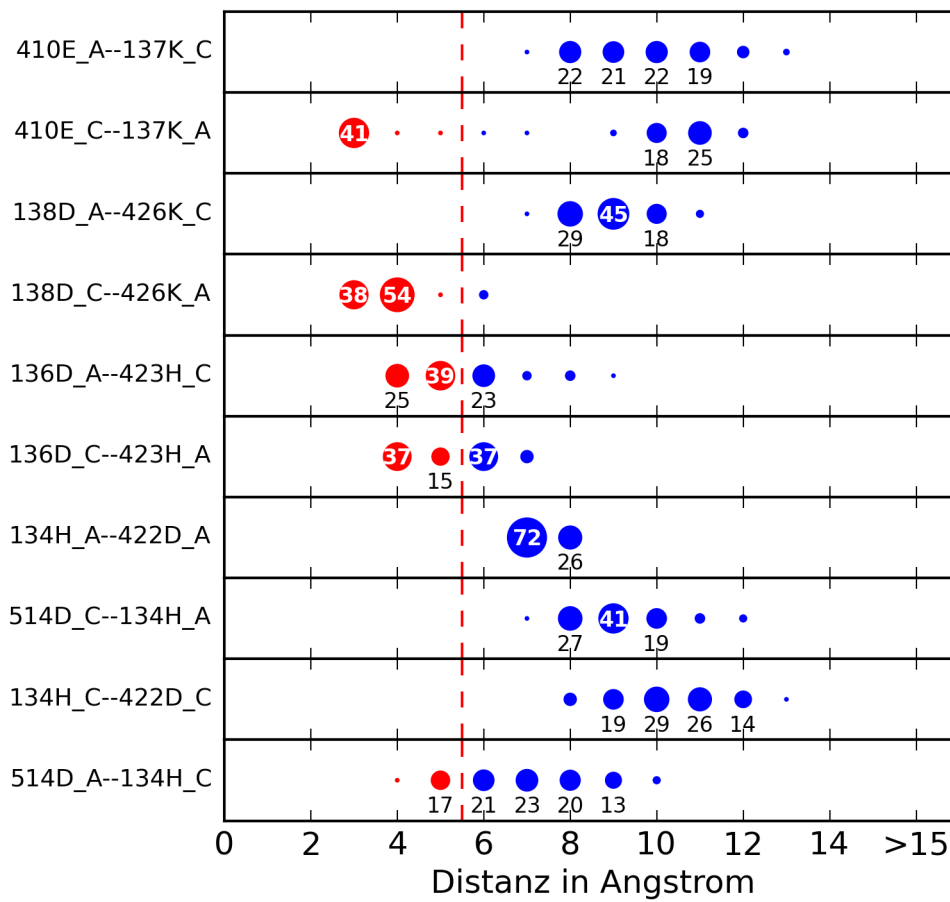


Abb. E.2: Entstanden aus den Graphen in Abb. E.1.

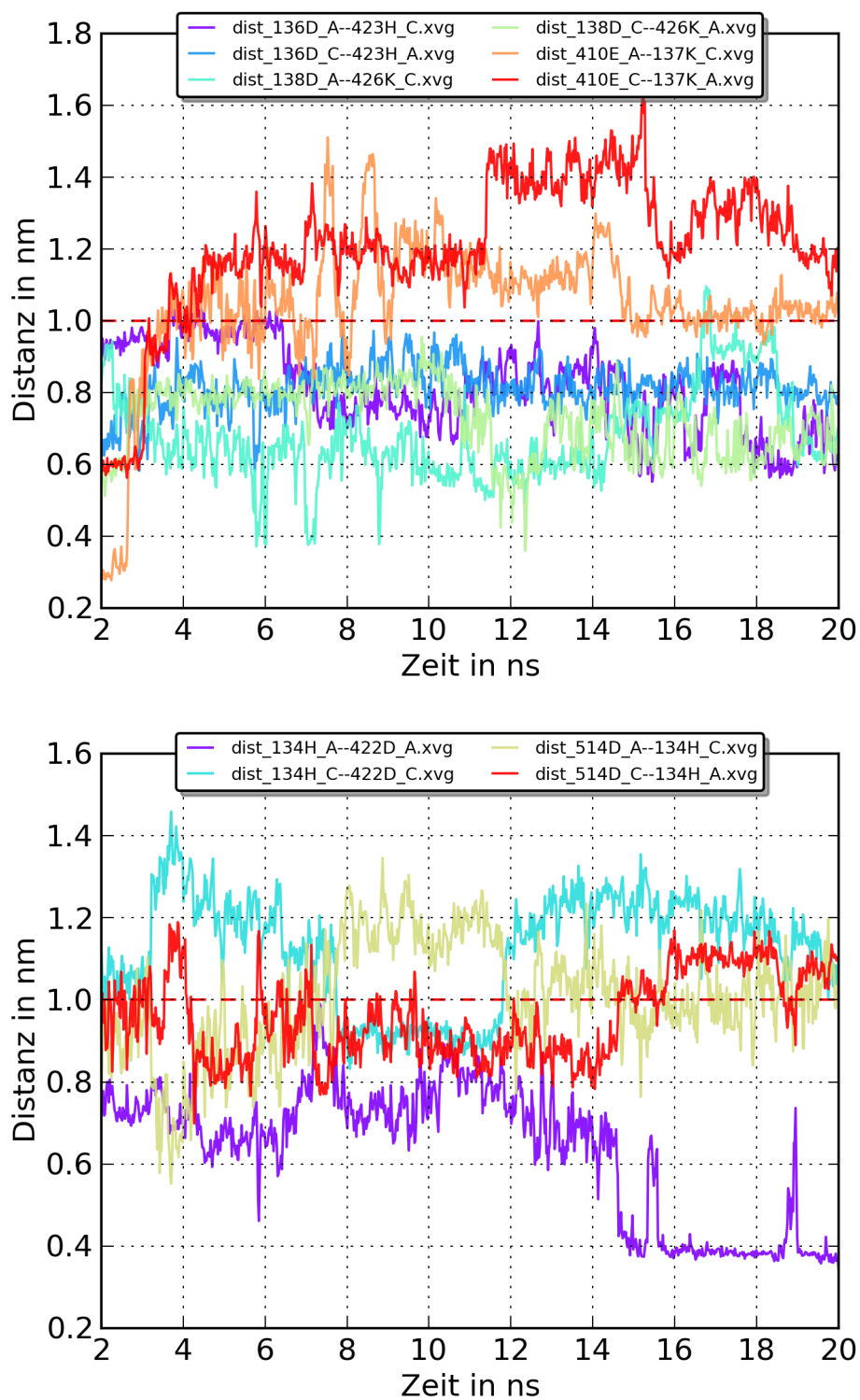


Abb. E.3: Aminosäurenabstände der 2. MD-Simulation im Puffer A mit 20 ns. Die Graphen wurden in das Scheibendiagramm in Abb. E.4 umgewandelt.

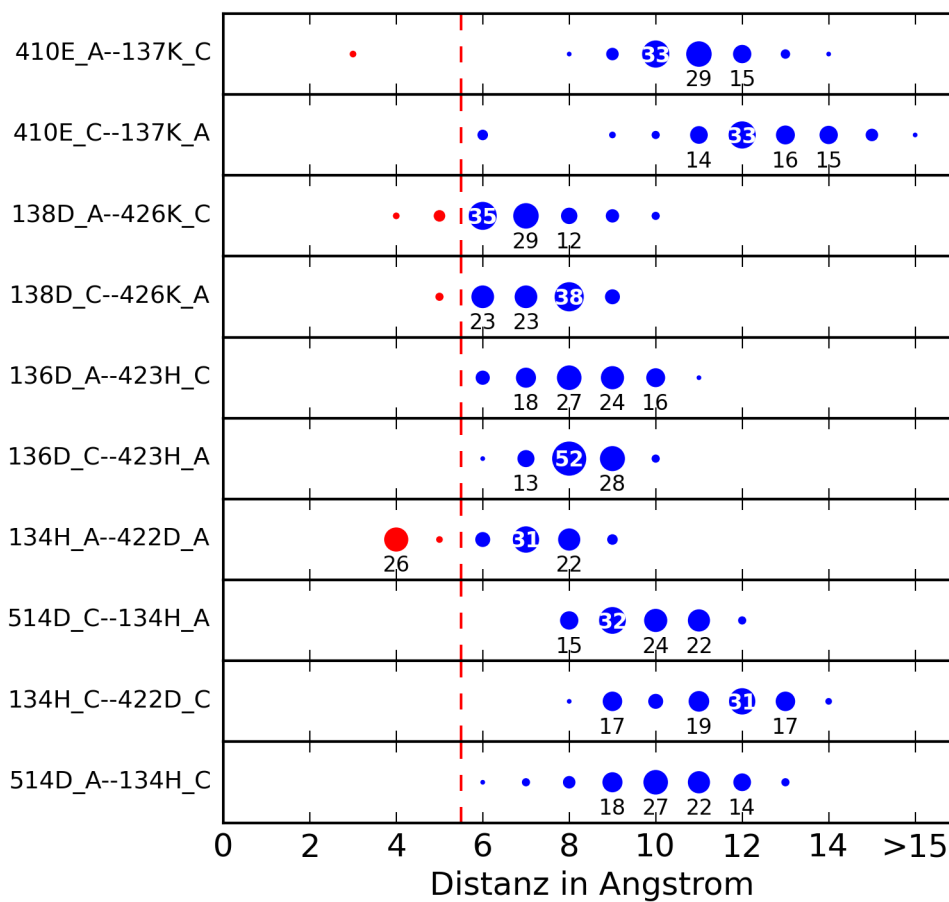
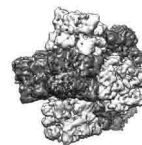


Abb. E.4: Entstanden aus den Graphen in Abb. E.3.

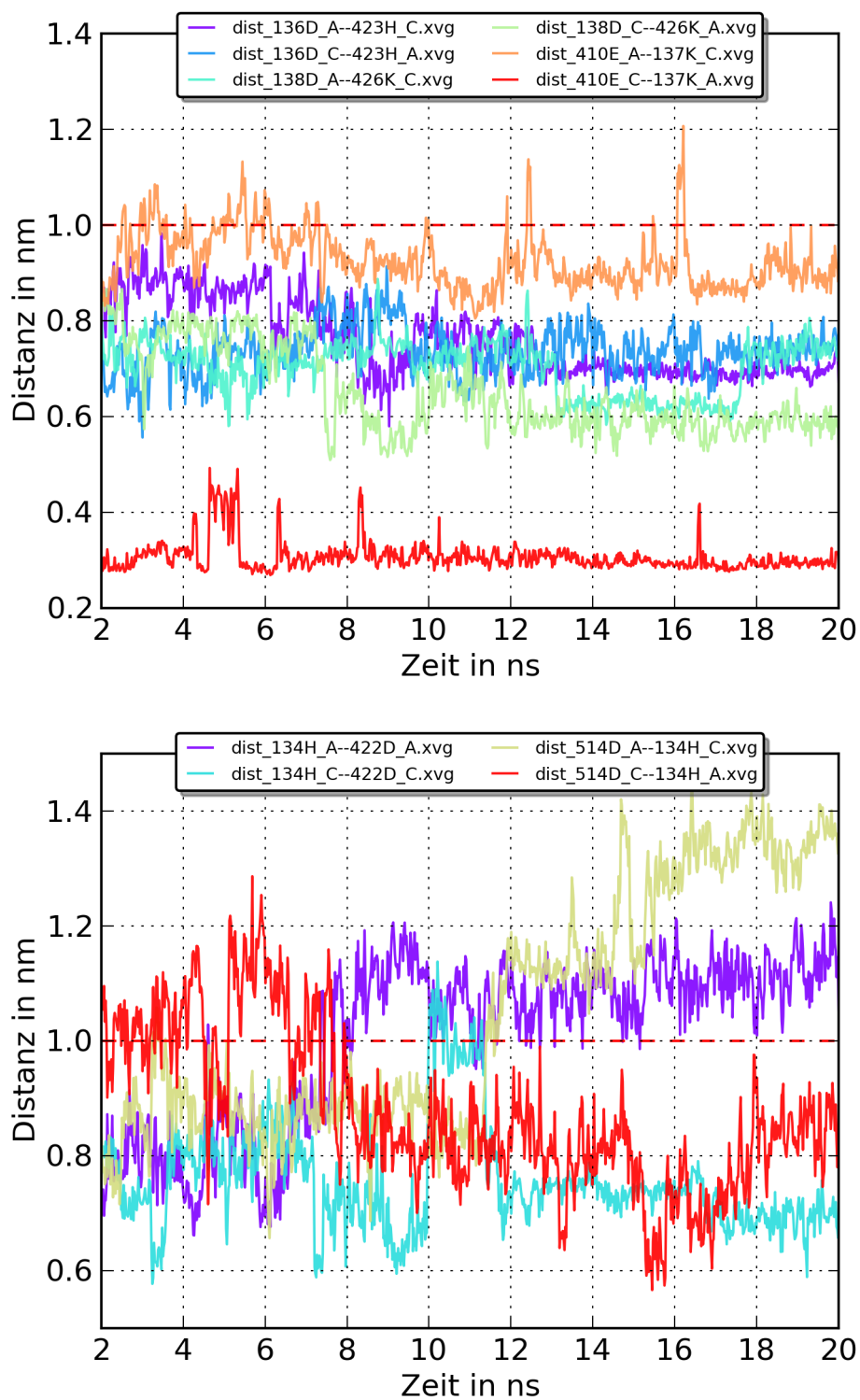


Abb. E.5: Aminosäureabstände der 3. MD-Simulation im Puffer A mit 20 ns. Die Graphen wurden in das Scheibendiagramm in Abb. E.6 umgewandelt.

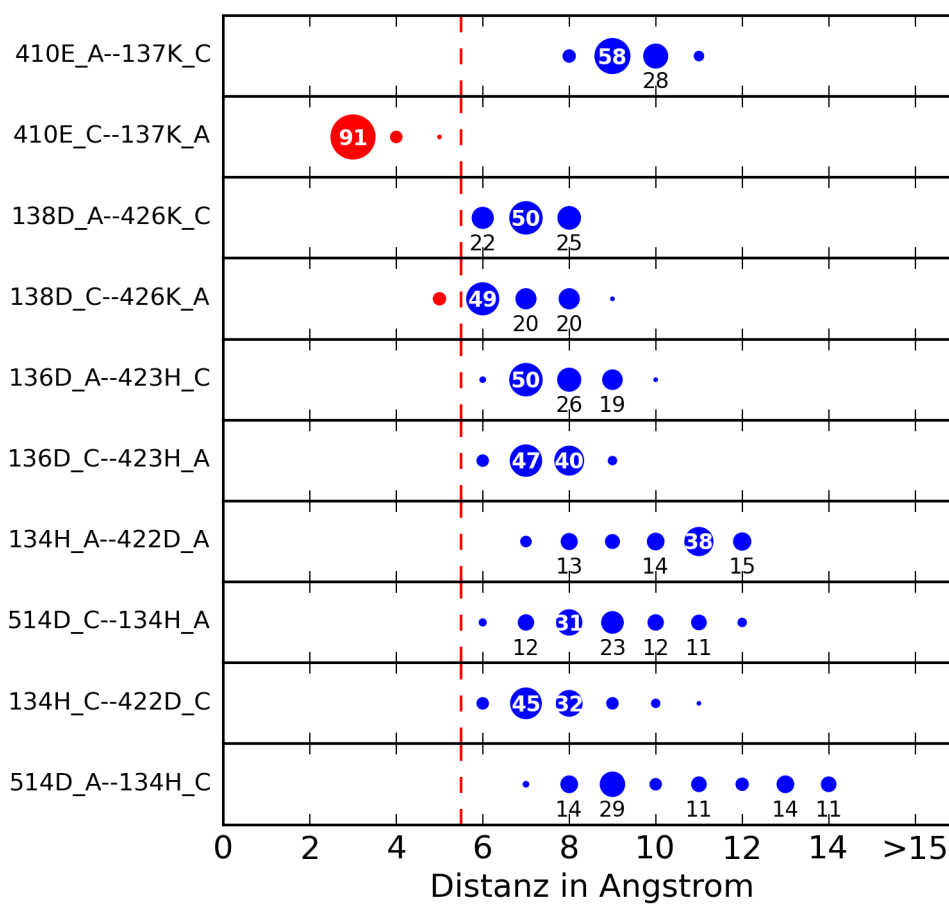
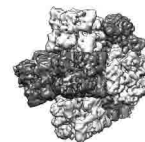


Abb. E.6: Entstanden aus den Graphen in Abb. E.5.

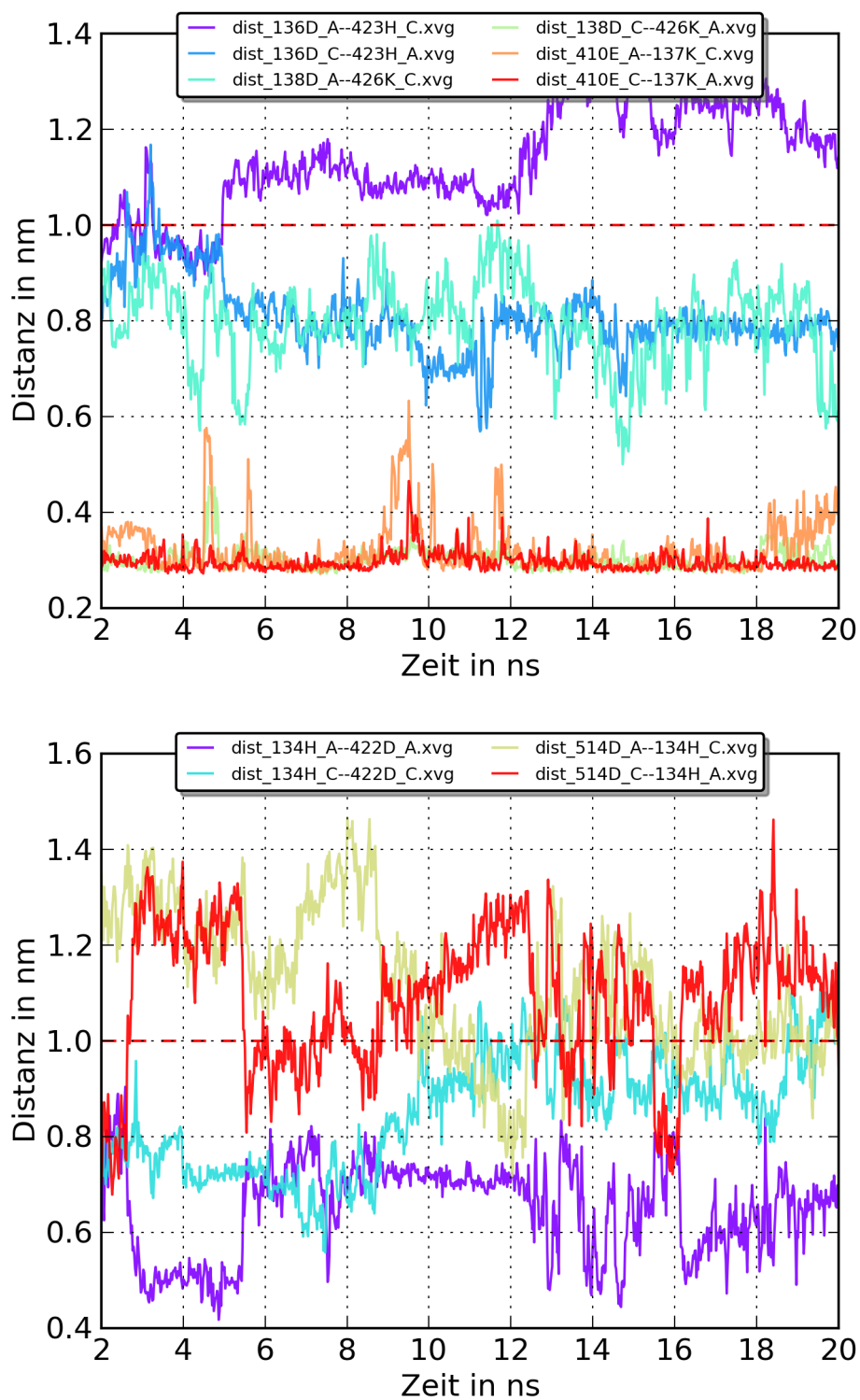


Abb. E.7: Aminosäurenabstände der 4. MD-Simulation im Puffer A mit 20 ns. Die Graphen wurden in das Scheibendiagramm in Abb. E.8 umgewandelt.

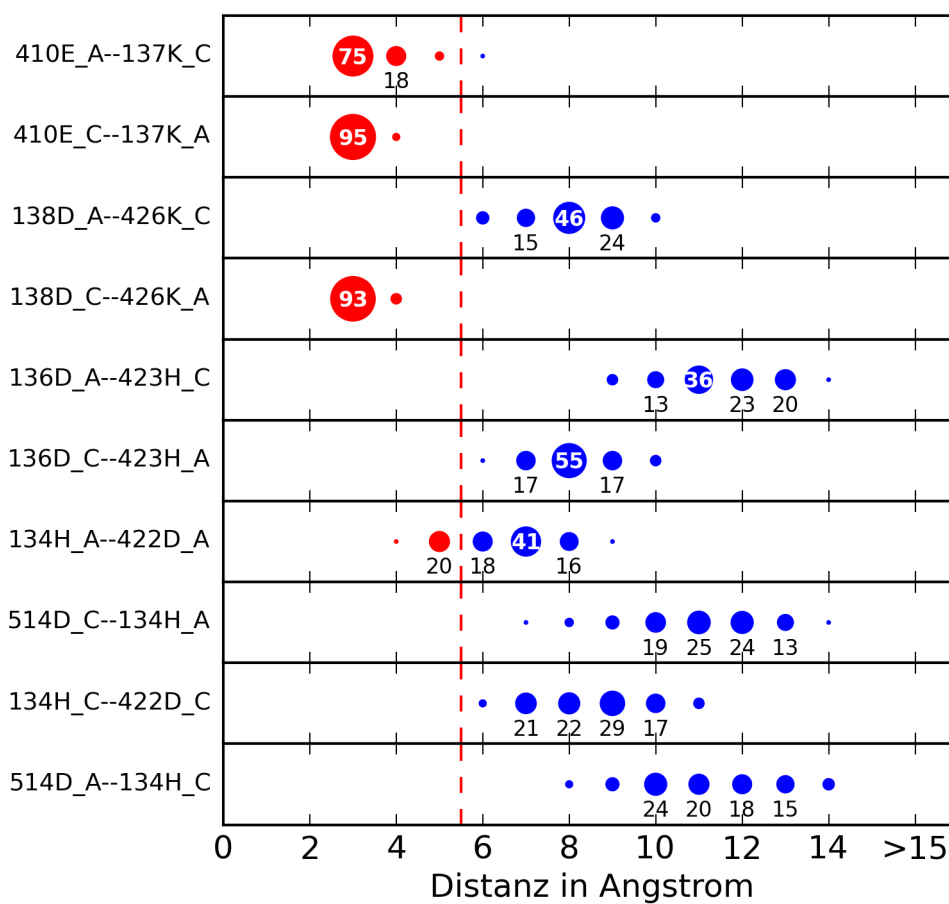


Abb. E.8: Entstanden aus den Graphen in Abb. E.7.

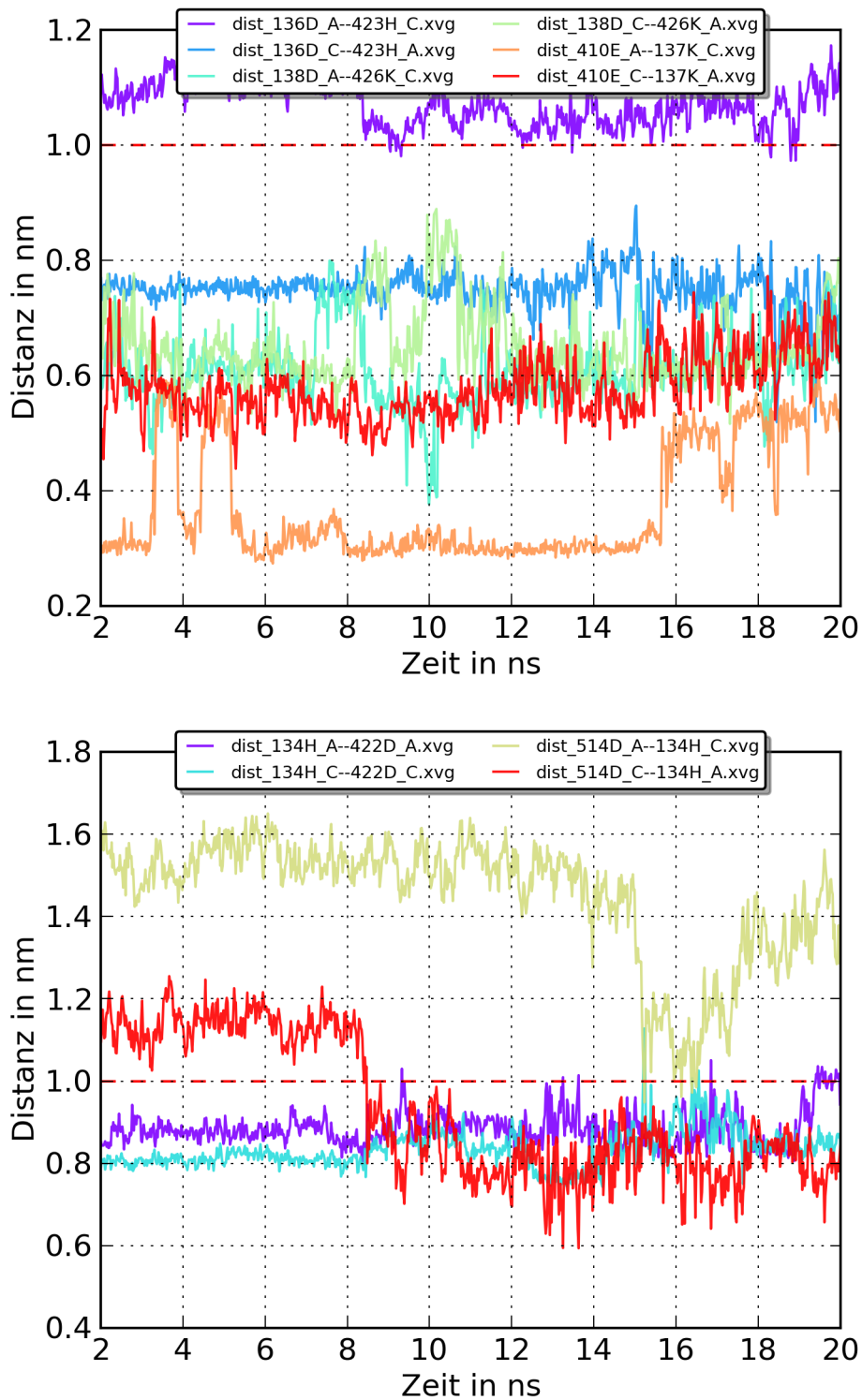


Abb. E.9: Aminosäurenabstände der 5. MD-Simulation im Puffer A mit 20 ns. Die Graphen wurden in das Scheibendiagramm in Abb. E.10 umgewandelt.

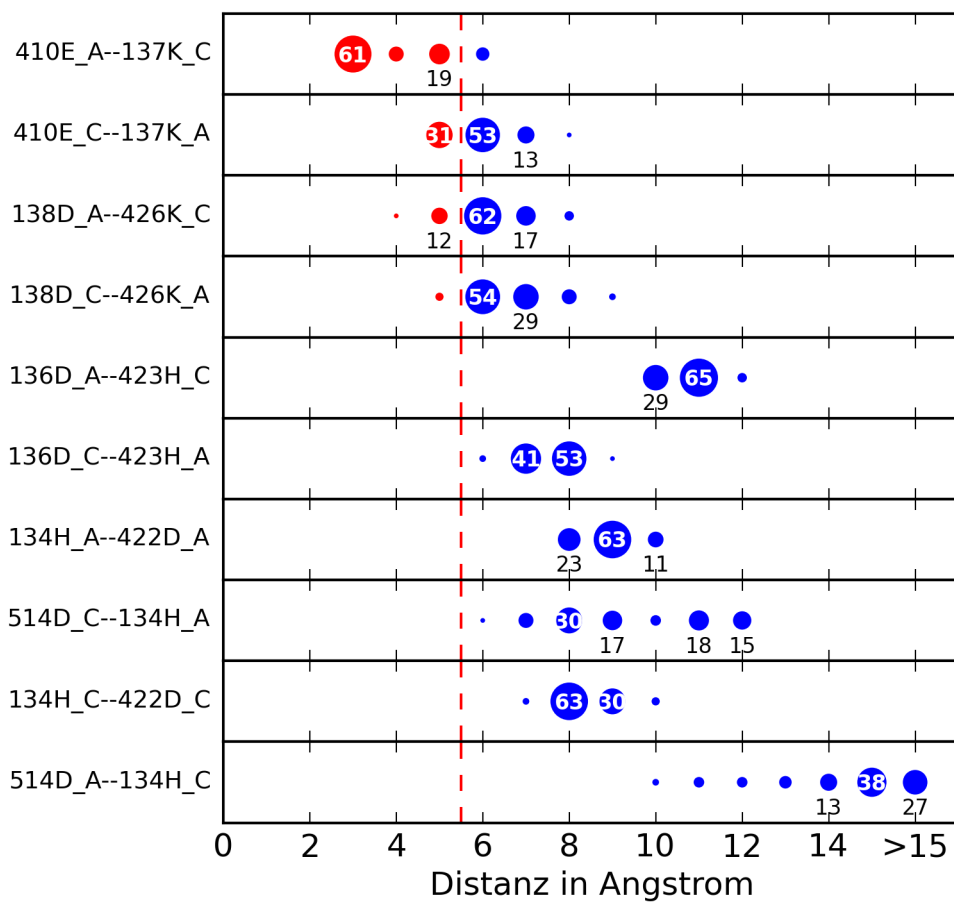
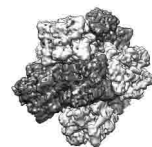


Abb. E.10: Entstanden aus den Graphen in Abb. E.9.

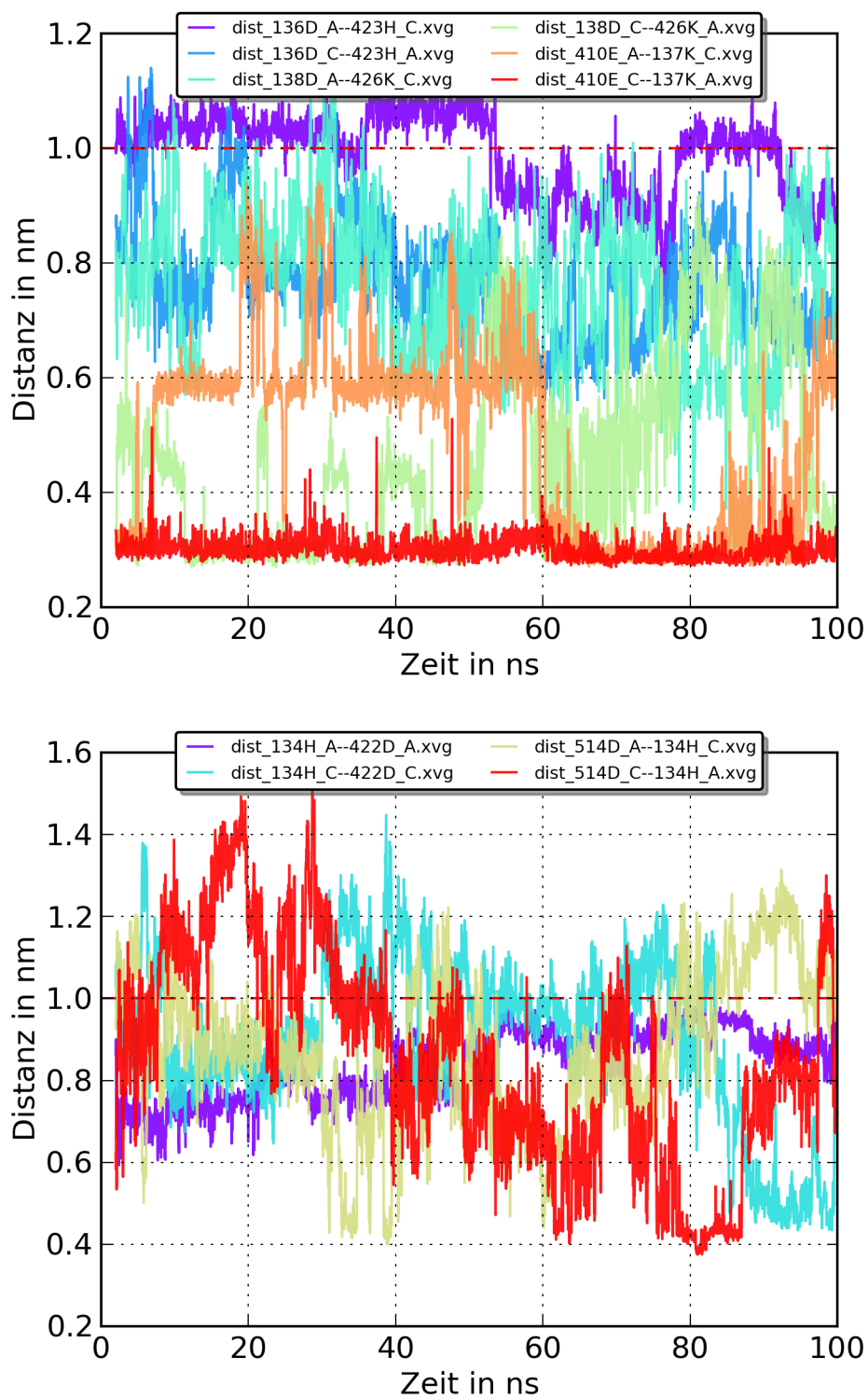


Abb. E.11: Aminosäurenabstände der 6. MD-Simulation im Puffer A mit 100 ns. Die Graphen wurden in das Scheibendiagramm in Abb. E.12 umgewandelt.

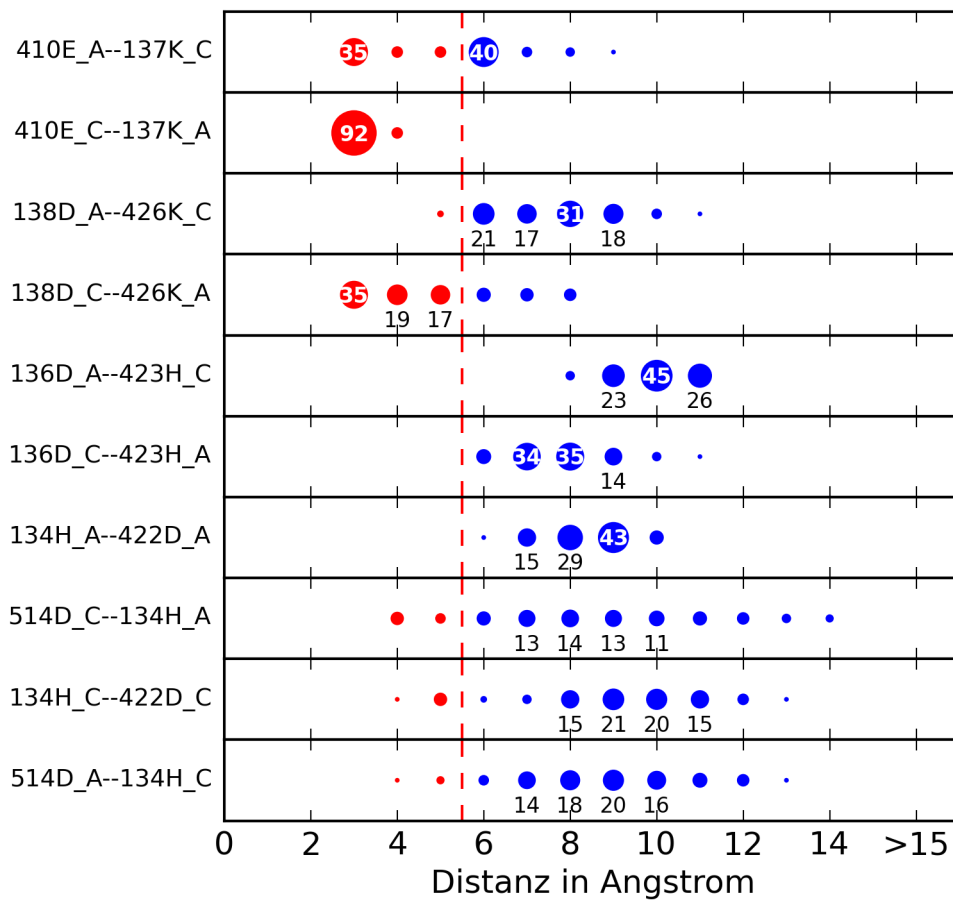
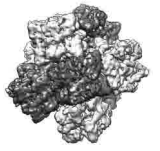


Abb. E.12: Entstanden aus den Graphen in Abb. E.11.

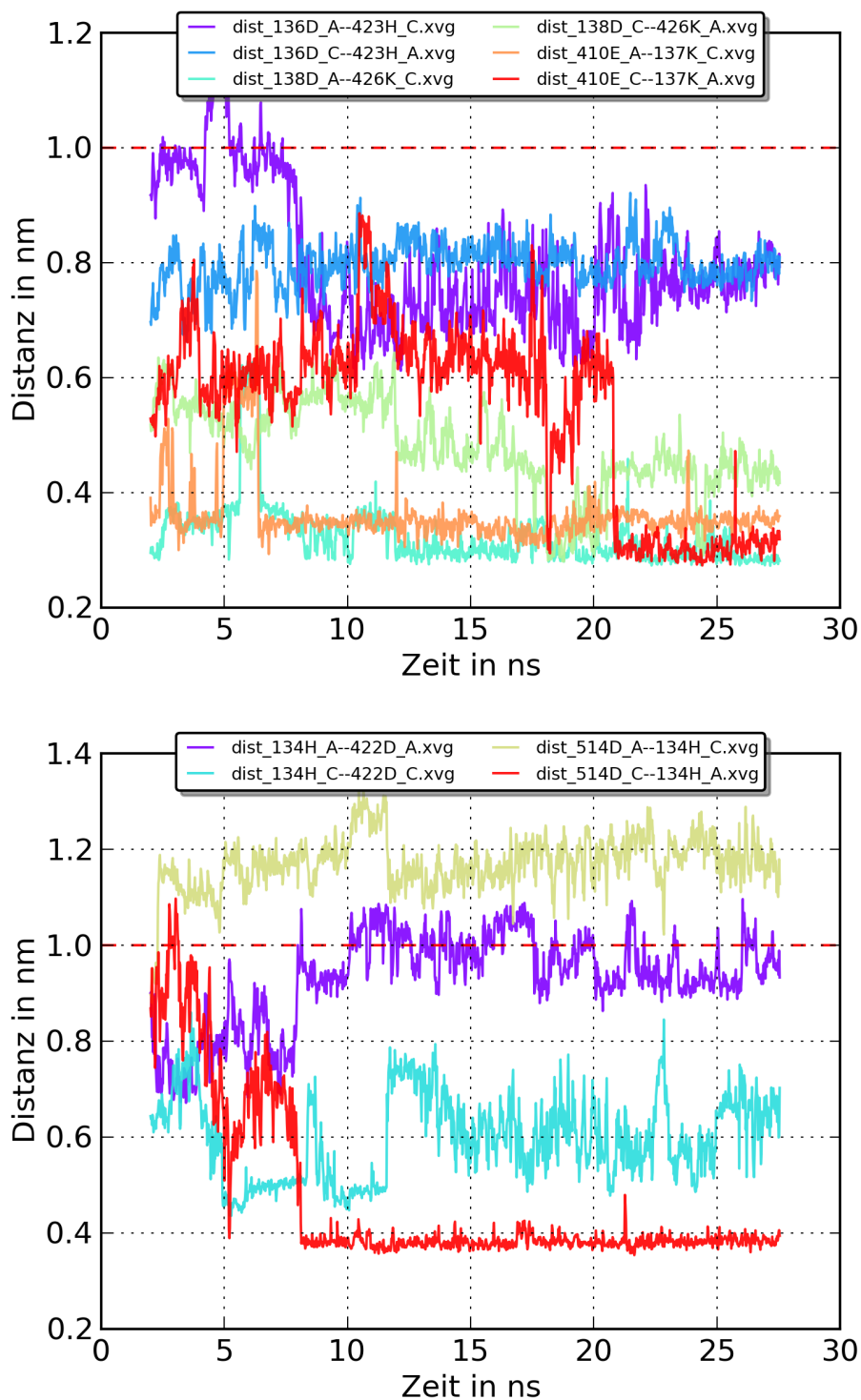


Abb. E.13: Aminosäurenabstände der 1. MD-Simulation im Puffer B mit 28 ns. Zum Zeitpunkt der Auswertung war die MD-Simulation mit einer Gesamtzeit von 100 ns noch nicht beendet. Die Graphen wurden in das Scheibendiagramm in Abb. E.14 umgewandelt.

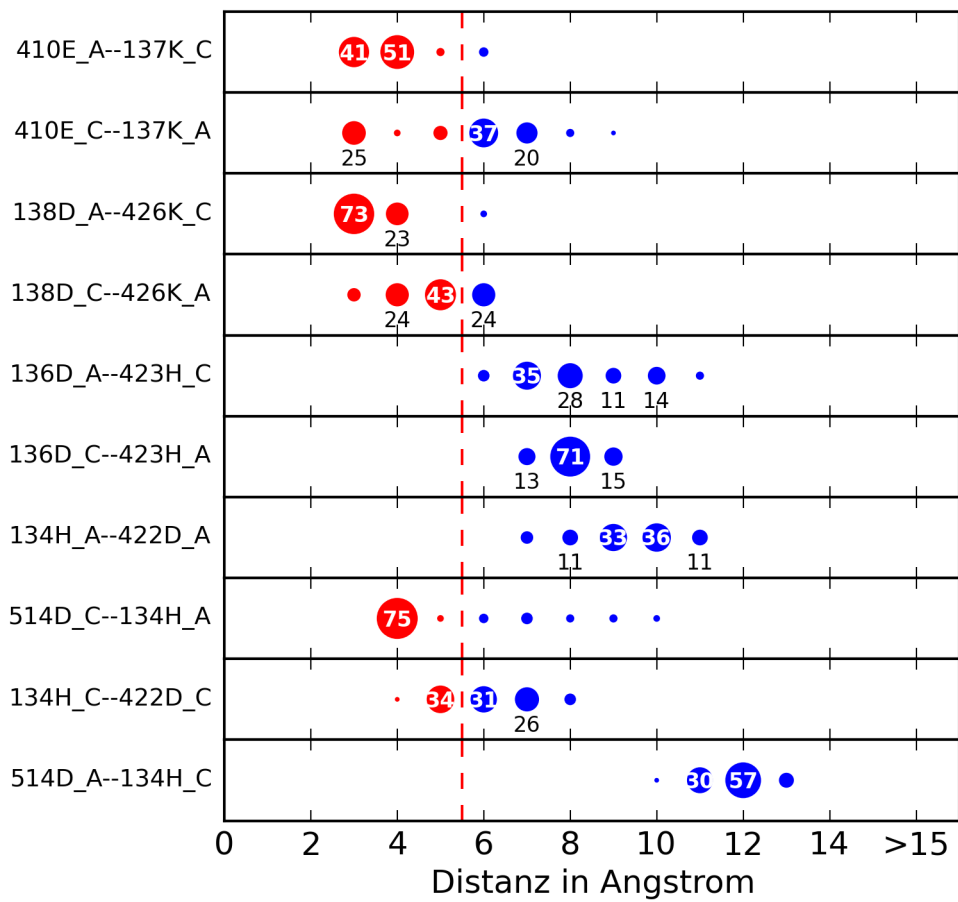
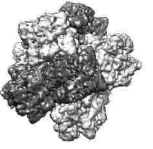
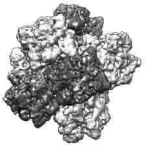


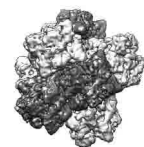
Abb. E.14: Entstanden aus den Graphen in Abb. E.13.



Literaturverzeichnis

- Ackers, G. K. (1970). Analytical gel chromatography of proteins. *Advances in protein chemistry*, 24:343–446.
- Adrian, M., Dubochet, J., and McDowell, A. (1984). Cryo-electron microscopy of viruses. *Nature*, 308.
- Aho, A. V., Kernighan, B. W., and Weinberger, P. J. (1979). Awk - a pattern scanning and processing language. *Software: Practice and Experience*, 9(4):267–279.
- Baker, M. L., Hryc, C. F., Zhang, Q., Wu, W., Jakana, J., Haase-Pettingell, C., Afonine, P. V., Adams, P. D., King, J. a., Jiang, W., and Chiu, W. (2013). Validated near-atomic resolution structure of bacteriophage epsilon15 derived from cryo-EM and modeling. *Proceedings of the National Academy of Sciences of the United States of America*, 110(30):12301–6.
- Bammes, B. E., Rochat, R. H., Jakana, J., Chen, D.-H., and Chiu, W. (2012). Direct electron detection yields cryo-EM reconstructions at resolutions beyond 3/4 Nyquist frequency. *Journal of structural biology*, 177(3):589–601.
- Beauchamp, K. a., Lin, Y.-S., Das, R., and Pande, V. S. (2012). Are Protein Force Fields Getting Better? A Systematic Benchmark on 524 Diverse NMR Measurements. *Journal of chemical theory and computation*, 8(4):1409–1414.
- Beck, K. (2002). *Test-driven Development: By Example*. Addison-Wesley Longman, Amsterdam.
- Beckmann, R. (1997). Alignment of Conduits for the Nascent Polypeptide Chain in the Ribosome-Sec61 Complex. *Science*, 278(5346):2123–2126.

- Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., and Bourne, P. E. (2000). The Protein Data Bank. *Nucleic acids research*, 28(1):235–42.
- Bijlholt, M. M., van Bruggen, E. F., and Bonaventura, J. (1979). Dissociation and reassembly of *Limulus polyphemus* hemocyanin. *European journal of biochemistry / FEBS*, 95(2):399–405.
- Böttcher, B., Wynne, S. A., and Crowther, R. A. (1997). Determination of the fold of the core protein of hepatitis B virus by electron cryomicroscopy. *Nature*, 386(6620):88–91.
- Bracewell, R. (1956). Strip Integration in Radio Astronomy. *Australian Journal of Physics*, 9(2):198.
- Brenner, S. and Horne, R. (1959). A negative staining method for high resolution electron microscopy of viruses. *Biochimica et Biophysica Acta*, 34:103–110.
- Brenowitz, M., Bonaventura, C., Bonaventure, J., and Gianazza, E. (1981). Subunit Composition of a High Molecular Weight Oligomer: *Limulus polyphemus* Hemocyanin. *Archives of Biochemistry and Biophysics*, 210(2):748–761.
- Carragher, B., Kisseberth, N., Kriegman, D., Milligan, R. a., Potter, C. S., Purokas, J., and Reilein, a. (2000). Leginon: an automated system for acquisition of images from vitreous ice specimens. *Journal of structural biology*, 132(1):33–45.
- Chen, J. Z. and Grigorieff, N. (2007). SIGNATURE: a single-particle selection system for molecular electron microscopy. *Journal of structural biology*, 157(1):168–73.
- Cluster Resources, I. (2009). *TORQUE Administrator's Guide*.
- Colbourne, J. K., Pfrender, M. E., Gilbert, D., Thomas, W. K., Tucker, A., Oakley, T. H., Tokishita, S., Aerts, A., Arnold, G. J., Basu, M. K., Bauer, D. J., Cáceres, C. E., Carmel, L., Casola, C., Choi, J.-H., Detter, J. C., Dong, Q., Dusheyko, S., Eads, B. D., Fröhlich, T., Geiler-Samerotte, K. a., Gerlach, D., Hatcher, P., Jogdeo, S., Krijgsveld, J., Kriventseva, E. V., Kültz, D., Laforsch,



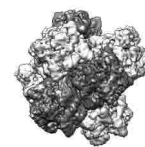
- C., Lindquist, E., Lopez, J., Manak, J. R., Muller, J., Pangilinan, J., Patwardhan, R. P., Pitluck, S., Pritham, E. J., Rechtsteiner, A., Rho, M., Rogozin, I. B., Sakarya, O., Salamov, A., Schaack, S., Shapiro, H., Shiga, Y., Skalitzky, C., Smith, Z., Souvorov, A., Sung, W., Tang, Z., Tsuchiya, D., Tu, H., Vos, H., Wang, M., Wolf, Y. I., Yamagata, H., Yamada, T., Ye, Y., Shaw, J. R., Andrews, J., Crease, T. J., Tang, H., Lucas, S. M., Robertson, H. M., Bork, P., Koonin, E. V., Zdobnov, E. M., Grigoriev, I. V., Lynch, M., and Boore, J. L. (2011). The ecoresponsive genome of *Daphnia pulex*. *Science (New York, N.Y.)*, 331(6017):555–61.
- Cong, Y., Zhang, Q., Woolford, D., Schweikardt, T., Khant, H., Dougherty, M., Ludtke, S. J., Chiu, W., and Decker, H. (2009). Structural mechanism of SDS-induced enzyme activity of scorpion hemocyanin revealed by electron cryomicroscopy. *Structure (London, England : 1993)*, 17(5):749–58.
- Conway, J. F., Cheng, N., Zlotnick, A., Wingfield, P. T., Stahl, S. J., and Steven, A. C. (1997). Visualization of a 4-helix bundle in the hepatitis B virus capsid by cryo-electron microscopy. *Nature*, 386(6620):91–4.
- Conway, J. F. and Steven, A. C. (1999). from Cryoelectron Micrographs to Subnanometer Resolution. 118:106–118.
- Cosslett, V. E., Camps, R. A., Saxton, W. O., Smith, D. J., Nixon, W. C., Ahmed, H., Catto, C. J. D., Cleaver, J. R. A., Smith, K. C. A., Timbs, A. E., Turner, P. W., and Ross, P. M. (1979). Atomic resolution with a 600-kV electron microscope. *Nature*, 281(5726):49–51.
- Crowther, R., Amos, L., Finch, J. T., DeRosier, D. J., and Klug, A. (1970a). Three Dimensional Reconstructions of Spherical Viruses by Fourier Synthesis from Electron Micrographs. *Nature*, 226.
- Crowther, R., DeRosier, D. J., and Klug, A. (1970b). The Reconstruction of a Three-Dimensional Structure from Projections and its Application to Electron Microscopy. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 317(1530):319–340.
- Daberkow, I., Herrmann, K.-H., Liu, L., Rau, W., and Tietz, H. (1996). Development and performance of a fast fibre-plate coupled CCD camera at medium

- energy and image processing system for electron holography. *Ultramicroscopy*, 64(1-4):35–48.
- Dangott, L. and Terwilliger, R. (1980). The subunit structure of *Daphnia pulex* hemoglobin. *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology*, 67(2):301–306.
- de Groot, B. L., van Aalten, D. M., Scheek, R. M., Amadei, a., Vriend, G., and Berendsen, H. J. (1997). Prediction of protein conformational freedom from distance constraints. *Proteins*, 29(2):240–51.
- Decker, H. and Föll, R. (2000). Temperature adaptation influences the aggregation state of hemocyanin from *Astacus leptodactylus*. *Comparative biochemistry and physiology. Part A, Molecular & integrative physiology*, 127(2):147–54.
- DeRosier, D. J. and Klug, A. (1968). Reconstruction of three dimensional structures from electron micrographs. *Nature*, 217(5124):130–4.
- Dewilde, S., Van Hauwaert, M. L., Peeters, K., Vanfleteren, J., and Moens, L. (1999). *Daphnia pulex* didomain hemoglobin: structure and evolution of polymeric hemoglobins and their coding genes. *Molecular biology and evolution*, 16(9):1208–18.
- Downing, K. H. and Glaeser, R. M. (2008). Restoration of weak phase-contrast images recorded with a high degree of defocus: the "twin image" problem associated with CTF correction. *Ultramicroscopy*, 108(9):921–8.
- Dubochet, J., Adrian, M., Chang, J. J., Homo, J. C., Lepault, J., McDowell, a. W., and Schultz, P. (1988). Cryo-electron microscopy of vitrified specimens. *Quarterly Reviews of Biophysics*, 21(2):129–228.
- Estrozi, L. F. and Navaza, J. (2010). Ab initio high-resolution single-particle 3D reconstructions: the symmetry adapted functions way. *Journal of structural biology*, 172(3):253–60.
- Eswar, N., Webb, B., Marti-Renom, M. a., Madhusudhan, M. S., Eramian, D., Shen, M.-Y., Pieper, U., and Sali, A. (2007). Comparative protein structure modeling using MODELLER. *Current protocols in protein science / editorial board, John E. Coligan ... [et al.]*, Chapter 2:Unit 2.9.



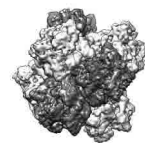
- Fernández, J. J., Luque, D., Castón, J. R., and Carrascosa, J. L. (2008). Sharpening high resolution information in single particle electron cryomicroscopy. *Journal of structural biology*, 164(1):170–5.
- Figura, M. and Gross, D. (2013). Die Qual der Wiki - Wahl Wikis für Wissensmanagement in Organisationen.
- Fox, H. M. (1948). The Haemoglobin of *Daphnia*. *Proceedings of the Royal Society B: Biological Sciences*, 135(879):195–212.
- Frank, J. (2002). Single-particle imaging of macromolecules by cryo-electron microscopy. *Annual review of biophysics and biomolecular structure*, 31:303–19.
- Freeman, S. and Pryce, N. (2009). *Growing Object-Oriented Software, Guided by Tests*.
- Frisch, A. I. (2003). *Unix System-Administration*. O'Reilly.
- Gerke, P., Börding, C., Zeis, B., and Paul, R. J. (2011). Adaptive haemoglobin gene control in *Daphnia pulex* at different oxygen and temperature conditions. *Comparative biochemistry and physiology. Part A, Molecular integrative physiology*, 159(1):56–65.
- Glaeser, R. (2004). Historical background: why is it important to improve automated particle selection methods? *Journal of Structural Biology*, 145(1-2):15–18.
- Glaeser, R. M. and Downing, K. H. (1992). Assessment of resolution in biological electron crystallography. *Ultramicroscopy*, 47(1-3):256–65.
- Goddard, T. D., Huang, C. C., and Ferrin, T. E. (2005). Software extensions to UCSF chimera for interactive visualization of large molecular assemblies. *Structure (London, England : 1993)*, 13(3):473–82.
- Goddard, T. D., Huang, C. C., and Ferrin, T. E. (2007). Visualizing density maps with UCSF Chimera. *Journal of structural biology*, 157(1):281–7.
- Gohr, A. (2013). DokuWiki Website. <http://www.dokuwiki.org>.

- Gohr, A., Hüttemann, D., Faust, D., and Fuchs-Kittowski, F. (2010). Quality check with DokuWiki for instant user feedback. *Proceedings of the 6th International Symposium on Wikis and Open Collaboration - WikiSym '10*, page 1.
- Grosse-Kunstleve, R. W., Sauter, N. K., Moriarty, N. W., Paul, D., and Grosse-kunstleve, R. W. (2002). The Computational Crystallography Toolbox: crystallographic algorithms in a reusable software framework. *Journal of Applied Crystallography*, pages 126–136.
- Harauz, G. and van Heel, M. (1986). Exact filters for general geometry three dimensional reconstruction. *Optik*, 78(4).
- Harris, J. R. and Agutter, P. (1970). A negative staining study of human erythrocyte ghosts and rat liver nuclear membranes. *Journal of ultrastructure research*, 33(3):219–32.
- Harris, J. R. and Horne, R. W. (1994). Negative staining: A brief assessment of current technical benefits, limitations and future possibilities. *Micron*, 25(1):5–13.
- Hartigan, J. A. (1975). *Clustering Algorithms*. John Wiley and Sons, Inc.
- Heiland, R., Mooney, S. D., Boverhof, J., Jackson, K., Swat, M., Balter, A., Christie, M., and Insley, J. (2007). Python for Scientific Gateways Development. *Digital Media Library*.
- Henderson, R. (1995). The potential and limitations of neutrons, electrons and X-rays for atomic resolution microscopy of unstained biological molecules. *Quarterly reviews of biophysics*, 28(2):171–193.
- Herold, H. (1991). *AWK und SED*. Addison-Wesley Longman, Amsterdam, 1. edition.
- Herold, H. (1999). *Linux-Unix-Shells. Bourne-Shell, Korn-Shell, C-Shell, bash, tcsh*. Addison-Wesley Longman, Amsterdam.
- Hess, B., Uppsala, S., and Lindahl, E. (2008). GROMACS 4 : Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *Journal of chemical theory and computation*, 4:435–447.



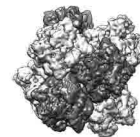
- Hess, B., Van Der Spoel, D., and Lindahl, E. (2013). *Gromacs User Manual Ver. 4.6.3*.
- Hohn, M., Tang, G., Goodyear, G., Baldwin, P. R., Huang, Z., Penczek, P. A., Yang, C., Glaeser, R. M., Adams, P. D., and Ludtke, S. J. (2007). SPARX, a new environment for Cryo-EM image processing. *Journal of structural biology*, 157(1):47–55.
- Horne, R. (1991). Early developments in the negative staining technique for electron microscopy. *Micron and Microscopica Acta*, 22(4):321–326.
- Humphrey, W., Dalke, A., and Schulten, K. (1996). VMD: Visual molecular dynamics. *Journal of Molecular Graphics*, 14(1):33–38.
- Hüttemann, D. and Riebesell, J. (2013). Wikimatrix Website. <http://www.wikimatrix.org>.
- Ilan, E., Weisselberg, E., and Daniel, E. (1982). Erythrocrucorin from the waterflea *Daphnia magna*. Quaternary structure and arrangement of subunits. *The Biochemical journal*, 207(2):297–303.
- Jeffrey, P. D., Shaw, D. C., and Treacy, G. B. (1978). Hemocyanin from the Australian freshwater crayfish *Cherax destructor*. Characterization of a dimeric subunit and its involvement in the formation of the 25S component. *Biochemistry*, 17(15):3078–84.
- Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983). Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics*, 79(2):926.
- Jorgensen, W. L., Jenson, C., and Haven, N. (1998). Carlo Simulations : Seeking Temperatures of Maximum Density. 19(10):1179–1186.
- Joyeux, L. and Penczek, P. A. (2002). Efficiency of 2D alignment methods. *Ultra-microscopy*, 92(2):33–46.
- Julenius, K., Mø lgaard, A., Gupta, R., and Brunak, S. r. (2005). Prediction, conservation analysis, and structural characterization of mammalian mucin-type O-glycosylation sites. *Glycobiology*, 15(2):153–64.

- Kendrew, J. C. and Parrish, R. G. (1957). The Crystal Structure of Myoglobin. III. Sperm-Whale Myoglobin. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 238(1214):305–324.
- Kisielowski, C., Freitag, B., Bischoff, M., van Lin, H., Lazar, S., Knippels, G., Tiemeijer, P., van der Stam, M., von Harrach, S., Stekelenburg, M., Haider, M., Uhlemann, S., Müller, H., Hartel, P., Kabius, B., Miller, D., Petrov, I., Olson, E. A., Donchev, T., Kenik, E. A., Lupini, A. R., Bentley, J., Pennycook, S. J., Anderson, I. M., Minor, A. M., Schmid, A. K., Duden, T., Radmilovic, V., Ramasse, Q. M., Watanabe, M., Erni, R., Stach, E. A., Denes, P., and Dahmen, U. (2008). Detection of single atoms and buried defects in three dimensions by aberration-corrected electron microscope with 0.5-Å information limit. *Microscopy and microanalysis : the official journal of Microscopy Society of America, Microbeam Analysis Society, Microscopical Society of Canada*, 14(5):469–77.
- Knoll, M. and Ruska, E. (1932). Das Elektronenmikroskop. *Zeitschrift für Physik*, 78(5-6):318–339.
- Kuwada, T., Hasegawa, T., Sato, S., Sato, I., Ishikawa, K., Takagi, T., and Shishikura, F. (2007). Crystal structures of two hemoglobin components from the midge larva *Prosilocerus akamusi* (Orthoclaadiinae, Diptera). *Gene*, 398(1-2):29–34.
- Lamkemeyer, T., Paul, R. J., Stöcker, W., Yiallourous, I., and Zeis, B. (2005). Macromolecular isoforms of *Daphnia magna* haemoglobin. *Biological chemistry*, 386(11):1087–96.
- Lamkemeyer, T., Zeis, B., Decker, H., Jaenicke, E., Waschbüsch, D., Gebauer, W., Markl, J., Meissner, U., Rousselot, M., Zal, F., Nicholson, G. J., and Paul, R. J. (2006). Molecular mass of macromolecules and subunits and the quaternary structure of hemoglobin from the microcrustacean *Daphnia magna*. *The FEBS journal*, 273(14):3393–410.
- Lamy, J., Lamy, J., Sizaret, P.-y., Billiald, P., Jollbs, P., Jollbs, J., Feldmann, R. J., and Bonaventura, J. (1983). Quaternary Structure of *Limulus polyphemus* Hemocyanin. *Biochemistry*, pages 5573–5583.



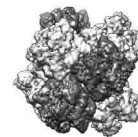
- Lander, G. C., Stagg, S. M., Voss, N. R., Cheng, A., Fellmann, D., Pulokas, J., Yoshioka, C., Irving, C., Mulder, A., Lau, P.-W., Lyumkis, D., Potter, C. S., and Carragher, B. (2009). Appion: An integrated, database-driven pipeline to facilitate EM image processing. *Journal of Structural Biology*, 166(1):95–102.
- Lawson, C. L., Baker, M. L., Best, C., Bi, C., Dougherty, M., Feng, P., van Ginkel, G., Devkota, B., Lagerstedt, I., Ludtke, S. J., Newman, R. H., Oldfield, T. J., Rees, I., Sahni, G., Sala, R., Velankar, S., Warren, J., Westbrook, J. D., Henrick, K., Kleywegt, G. J., Berman, H. M., and Chiu, W. (2011). EMDDataBank.org: unified data resource for CryoEM. *Nucleic acids research*, 39(Database issue):D456–64.
- Li, X., Mooney, P., Zheng, S., Booth, C. R., Braunfeld, M. B., Gubbens, S., Agard, D. A., and Cheng, Y. (2013). Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nature methods*, 10(6):584–90.
- Lindorff-Larsen, K., Maragakis, P., Piana, S., Eastwood, M. P., Dror, R. O., and Shaw, D. E. (2012). Systematic validation of protein force fields against experimental data. *PloS one*, 7(2):e32131.
- Loman, N. and Watson, M. (2013). So you want to be a computational biologist? *Nature biotechnology*, 31(11):996–8.
- Ludtke, S. J. (2012). EMAN2/BoxSize Website. <http://blake.bcm.edu/emanwiki/EMAN2/BoxSize>.
- Ludtke, S. J., Baldwin, P. R., and Chiu, W. (1999). EMAN: semiautomated software for high-resolution single-particle reconstructions. *Journal of structural biology*, 128(1):82–97.
- Mackall, M. (2006). Towards a Better SCM: Revlog and Mercurial. *Linux Symposium*, 2:91–98.
- Mallick, S. (2004). Detecting particles in cryo-EM micrographs using learned features. *Journal of Structural Biology*, 145(1-2):52–62.
- Mallick, S. P., Carragher, B., Potter, C. S., and Kriegman, D. J. (2005). ACE: automated CTF estimation. *Ultramicroscopy*, 104(1):8–29.

- Markl, J. (2013). Evolution of molluscan hemocyanin structures. *Biochimica et Biophysica Acta (BBA)-Proteins and . . .*, pages 26–31.
- Markl, J. and Decker, H. (1992). Molecular Structure of the Arthropod Hemocyanins. In *Blood and Tissue Oxygen Carriers*, volume 13.
- Markl, J., Kempter, B., Linzen, B., Bijlholt, M. M., and van Bruggen, E. F. (1981). Hemocyanins in spiders, XVI[1]. Subunit topography and a model of the quaternary structure of Eurypelma hemocyanin. *Hoppe-Seyleyler's Zeitschrift für physiologische Chemie*, 362(12):1631–41.
- Markl, J., Markl, A., Schartau, W., and Linzen, B. (1979). Subunit Heterogeneity in Arthropod Hemocyanins : I . Chelicerata *. *Journal of Comparative Physiology B*, 292:283–292.
- Markl, J., Moeller, A., Martin, A. G., Rheinbay, J., Gebauer, W., and Depoix, F. (2009). 10-A cryoEM structure and molecular model of the Myriapod (Scutigera) 6x6mer hemocyanin: understanding a giant oxygen transport protein. *Journal of molecular biology*, 392(2):362–80.
- Martin, A. G., Depoix, F., Stohr, M., Meissner, U., Hagner-Holler, S., Hammouti, K., Burmester, T., Heyd, J., Wriggers, W., and Markl, J. (2007). *Limulus polyphemus* hemocyanin: 10 A cryo-EM structure, sequence analysis, molecular modelling and rigid-body fitting reveal the interfaces between the eight hexamers. *Journal of molecular biology*, 366(4):1332–50.
- Martin, R. C. (2008). *Clean Code: A Handbook of Agile Software Craftsmanship*. Prentice Hall International.
- McDonald, G. D., Davidson, L., and Kitto, G. B. (1992). Amino acid sequence of the coelomic C globin from the sea cucumber *Caudina (Molpadia) arenicola*. *Journal of protein chemistry*, 11(1):29–37.
- Mindell, J. a. and Grigorieff, N. (2003). Accurate determination of local defocus and specimen tilt in electron microscopy. *Journal of Structural Biology*, 142(3):334–347.



- Mičetić, I., Losasso, C., Muro, P. D., Tognon, G., Benedetti, P., and Beltramini, M. (2010). Solution structures of 2x6-meric and 4x6-meric hemocyanins of crustaceans *Carcinus aestuarii*, *Squilla mantis* and *Upogebia pusilla*. *Journal of structural biology*, 171(1):1–10.
- Mouche, F., Boisset, N., and Penczek, P. A. (2001). *Lumbricus terrestris* hemoglobin—the architecture of linker chains and structural variation of the central toroid. *Journal of structural biology*, 133(2-3):176–92.
- Orlova, E. V., Dube, P., Harris, J. R., Beckman, E., Zemlin, F., Markl, J., and van Heel, M. (1997). Structure of keyhole limpet hemocyanin type 1 (KLH1) at 15 Å resolution by electron cryomicroscopy and angular reconstitution. *Journal of molecular biology*, 271(3):417–37.
- PDB-Website (2013). PDB Statistics Website. http://www.pdb.org/pdb/static.do?p=general_information/pdb_statistics/index.html.
- Peeters, K., Mertens, J., Hebert, P., and Moens, L. (1990). The globin composition of *Daphnia pulex* hemoglobin and the comparison of the amino acid composition of invertebrate hemoglobins. *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology*, 97(2):369–381.
- Penczek, P. A. (2010a). Image Restoration in Cryo-Electron Microscopy. *Methods in Enzymology, Vol 482: Cryo-Em, Part B: 3-D Reconstruction*, 482:35–72.
- Penczek, P. A. (2010b). Resolution Measures in Molecular Electron Microscopy. *Methods in Enzymology, Vol 482: Cryo-Em, Part B: 3-D Reconstruction*, 482:73–100.
- Penczek, P. A., Radermacher, M., and Frank, J. (1992). Three-dimensional reconstruction of single particles embedded in ice. *Ultramicroscopy*, 40(1):33–53.
- Penczek, P. a., Renka, R., and Schomberg, H. (2004). Gridding-based direct Fourier inversion of the three-dimensional ray transform. *Journal of the Optical Society of America A*, 21(4):499.
- Penczek, P. A., Zhu, J., and Frank, J. (1996). A common-lines based method for determining orientations for $N > 3$ particle projections simultaneously. *Ultramicroscopy*, 63(3-4):205–218.

- Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. (2004). UCSF Chimera—a visualization system for exploratory research and analysis. *Journal of computational chemistry*, 25(13):1605–12.
- Potter, C. S., Chu, H., Frey, B., Green, C., Kisseberth, N., Madden, T. J., Miller, K. L., Nahrstedt, K., Pulokas, J., Reilein, a., Tchong, D., Weber, D., and Carragher, B. (1999). Leginon: a system for fully automated acquisition of 1000 electron micrographs a day. *Ultramicroscopy*, 77(3-4):153–61.
- Riverbank-Computing (2013). PyQt Website. <http://www.riverbankcomputing.co.uk/software/pyqt/license>.
- Roseman, a. (2004). FindEM—a fast, efficient program for automatic selection of particles from electron micrographs. *Journal of Structural Biology*, 145(1-2):91–99.
- Rosenthal, P. B. and Henderson, R. (2003). Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. *Journal of Molecular Biology*, 333:721–745.
- Rousselot, M., Jaenicke, E., Lamkemeyer, T., Harris, J. R., and Pirow, R. (2006). Native and subunit molecular mass and quaternary structure of the hemoglobin from the primitive branchiopod crustacean *Triops cancriformis*. *The FEBS journal*, 273(17):4055–71.
- Schatz, M. (1992). Invariante Klassifizierung elektronenmikroskopischer Aufnahmen von eiseingebetteten biologischen Makromolekülen.
- Schatz, M., Orlova, E. V., Dube, P., Jäger, J., and van Heel, M. (1995). Structure of *Lumbricus terrestris* hemoglobin at 30 Å resolution determined using angular reconstitution. *Journal of structural biology*, 114(1):28–40.
- Scheres, S. H. W. (2012). RELION: Implementation of a Bayesian approach to cryo-EM structure determination. *Journal of structural biology*, 180(3):519–530.
- Scheres, S. H. W. (2013). Single-particle processing in RELION.



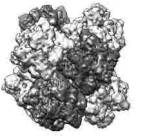
- Schröder, G. F., Brunger, A. T., and Levitt, M. (2007). Combining efficient conformational sampling with a deformable elastic network model facilitates structure refinement at low resolution. *Structure (London, England : 1993)*, 15(12):1630–41.
- Spindler, K.-D., Hennecke, R., and Gellissen, G. (1992). Protein Production and the Molting Cycle in the Crayfish *Astacus leptodactylus*. *General and Comparative Endocrinology*, 253:248–253.
- Stöcker, W., Raeder, U., Bijlholt, M. M. C., Wichertjes, T., Bruggen, E. F. J., and Markl, J. (1988). The quaternary structure of four crustacean two-hexameric hemocyanins: immunocorrelation, stoichiometry, reassembly and topology of individual subunits. *Journal of Comparative Physiology B*, 158(3):271–289.
- Sugano, H. and Hoshi, T. (1971). Purification and properties of blood hemoglobin from the fresh-water cladocera, *Moina macrocopa* and *Daphnia magna*. *Biochimica et Biophysica Acta (BBA) - Protein Structure*, 229(2):349–358.
- Summerfield, M. (2007). *Rapid GUI Programming with Python and Qt: The Definitive Guide to PyQt Programming*.
- Sutmann, G. (2002). *Classical Molecular Dynamics*, volume 10.
- Svedberg, T. and Eriksson-Quensel, I.-B. (1934). The Molecular Weight of Erythrocyrin. II. *Journal of the American Chemical Society*, 56(8):1700–1706.
- Talmon, Y. (1983). Staining and drying-induced artifacts in electron microscopy of surfactant dispersions. *Journal of Colloid and Interface Science*, 93(2):366–382.
- Tang, G., Peng, L., Baldwin, P. R., Mann, D. S., Jiang, W., Rees, I., and Ludtke, S. J. (2007). EMAN2: an extensible image processing suite for electron microscopy. *Journal of structural biology*, 157(1):38–46.
- Tang, G., Peng, L., Baldwin, P. R., Mann, D. S., Jiang, W., Rees, I., and Ludtke, S. J. (2012). e2eulerexplor Website. <http://blake.bcm.edu/emanwiki/EMAN2/Programs/e2eulerexplor>.

- Taveau, J. C., Boisset, N., Lamy, J., Lambert, O., and Lamy, J. N. (1997). Three-dimensional reconstruction of *Limulus polyphemus* hemocyanin from cryoelectron microscopy. *Journal of molecular biology*, 266(5):1002–15.
- Thomas, J. and Gemming, T. (2013). *Analytische Transmissionselektronenmikroskopie*. Springer Vienna, Vienna.
- Thon, F. (1966). Zur Defokussierungsabhängigkeit des Phasenkontrastes bei der elektronenmikroskopischen Abbildung. *Zeitschrift Naturforschung*, 21:476.
- Van Der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A. E., and Berendsen, H. J. C. (2005). GROMACS: fast, flexible, and free. *Journal of computational chemistry*, 26(16):1701–18.
- van Heel, M. (1987). Angular Reconstruction: A Posteriori Assignment of Projection Directions for 3D Reconstruction. *Ultramicroscopy*, 21:111–124.
- van Heel, M., Harauz, G., Orlova, E. V., Schmidt, R., and Schatz, M. (1996). A new generation of the IMAGIC image processing system. *Journal of structural biology*, 116(1):17–24.
- van Heel, M. and Keegstra, W. (1981). IMAGIC: A fast, flexible and friendly image analysis software system. *Ultramicroscopy*, 7:113–130.
- van Heel, M. and Schatz, M. (2005). Fourier shell correlation threshold criteria. *Journal of structural biology*, 151(3):250–62.
- van Rossum, G. and Drake, F. L. (2002). Python Reference Manual. Technical report.
- Veesler, D., Campbell, M. G., Cheng, A., Fu, C.-Y., Murez, Z., Johnson, J. E., Potter, C. S., and Carragher, B. (2013). Maximizing the potential of electron cryomicroscopy data collected using direct detectors. *Journal of structural biology*.
- Volbeda, a. and Hol, W. G. (1989). Crystal structure of hexameric haemocyanin from *Panulirus interruptus* refined at 3.2 Å resolution. *Journal of molecular biology*, 209(2):249–79.



- Voss, N., Yoshioka, C., Radermacher, M., Potter, C., and Carragher, B. (2009). DoG Picker and TiltPicker: Software tools to facilitate particle selection in single particle electron microscopy. *Journal of Structural Biology*, 166(2):205–213.
- Wang, Z. and Schröder, G. F. (2012). Real-space refinement with DireX: from global fitting to side-chain improvements. *Biopolymers*, 97(9):687–97.
- Weber, R. E. and Vinogradov, S. N. (2001). Nonvertebrate hemoglobins: functions and molecular adaptations. *Physiological reviews*, 81(2):569–628.
- Williams, D. B. and Carter, C. B. (2009). *Transmission Electron Microscope*.
- Wolf, G. H., Smet, J., and Declair, W. (1983). Oxygen binding properties of hemoglobins from *Daphnia pulex*. *Comparative Biochemistry and Physiology*, 75(2):261–265.
- Woolfson, M. M. (1970). *An introduction to X-ray crystallography*. Cambridge University Press.
- Yang, Z., Fang, J., Chittuluru, J., Asturias, F. J., and Penczek, P. A. (2012a). Iterative stable alignment and clustering of 2D transmission electron microscope images. *Structure (London, England : 1993)*, 20(2):237–47.
- Yang, Z., Lasker, K., Schneidman-Duhovny, D., Webb, B., Huang, C. C., Pettersen, E. F., Goddard, T. D., Meng, E. C., Sali, A., and Ferrin, T. E. (2012b). UCSF Chimera, MODELLER, and IMP: an integrated modeling system. *Journal of structural biology*, 179(3):269–78.
- Yoshioka, C., Carragher, B., and Potter, C. S. (2010). Cryomesh: a new substrate for cryo-electron microscopy. *Microscopy and microanalysis : the official journal of Microscopy Society of America, Microbeam Analysis Society, Microscopical Society of Canada*, 16(1):43–53.
- Yu, X., Ge, P., Jiang, J., Atanasov, I., and Zhou, Z. H. (2011). Atomic model of CPV reveals the mechanism used by this single-shelled virus to economically carry out functions conserved in multishelled reoviruses. *Structure (London, England : 1993)*, 19(5):652–61.

- Zeis, B., Becher, B., Lamkemeyer, T., Rolf, S., Pirow, R., and Paul, R. J. (2003). The process of hypoxic induction of *Daphnia magna* hemoglobin: subunit composition and functional properties. *Comparative biochemistry and physiology. Part B, Biochemistry & molecular biology*, 134(2):243–52.
- Zeis, B., Becker, D., Gerke, P., Koch, M., and Paul, R. J. (2013). Hypoxia-inducible haemoglobins of *Daphnia pulex* and their role in the response to acute and chronic temperature increase. *Biochimica et biophysica acta*, 1834(9):1704–10.
- Zernike, F. (1953). Wie ich den Phasenkontrast entdeckte. Technical report.
- Zernike, F. (1955). How I Discovered Phase Contrast. Technical Report 3141.
- Zhang, W., Kimmel, M., Spahn, C. M. T., and Penczek, P. a. (2008). Heterogeneity of large macromolecular complexes revealed by 3D cryo-EM variance analysis. *Structure (London, England : 1993)*, 16(12):1770–6.
- Zhang, X., Jin, L., Fang, Q., Hui, W. H., and Zhou, Z. H. (2010). 3.3 Å cryo-EM structure of a nonenveloped virus reveals a priming mechanism for cell entry. *Cell*, 141(3):472–82.
- Zhou, Z. H. (2008). Towards atomic resolution structural determination by single-particle cryo-electron microscopy. *Current opinion in structural biology*, 18:218–228.
- Zhu, J., Penczek, P. A., Schröder, R., and Frank, J. (1997). Three-dimensional reconstruction with contrast transfer function correction from energy-filtered cryoelectron micrographs: procedure and application to the 70S *Escherichia coli* ribosome. *Journal of structural biology*, 118(3):197–219.
- Ziade, T. (2008). *Expert Python Programming*.



Danksagung

Ich danke...

Erklärung und Copyright

Erklärung

Ich versichere hiermit, die vorliegende Arbeit selbstständig und nur mit den angegebenen Hilfsmitteln angefertigt zu haben.

Mainz, 24. April 2014

Ort, Datum

Mario Dejung

