

# Konformations- und Konfigurationshyperflächen – Generierung, Eigenschaften und Darstellung



## **Dissertation**

zur Erlangung des Doktorgrades der  
Mathematisch-Naturwissenschaftlichen Fakultät der  
Christian-Albrechts-Universität zu Kiel

vorgelegt von

**Felix Johannes Köhler**

geboren in Idar Oberstein

Kiel 2008

1. Gutachter: Prof. Dr. R. Herges
2. Gutachter: Prof. Dr. A. Krüger

Tag der Disputation: 03.06.2008

Zum Druck genehmigt: 03.06.2008

## Eidesstattliche Erklärung

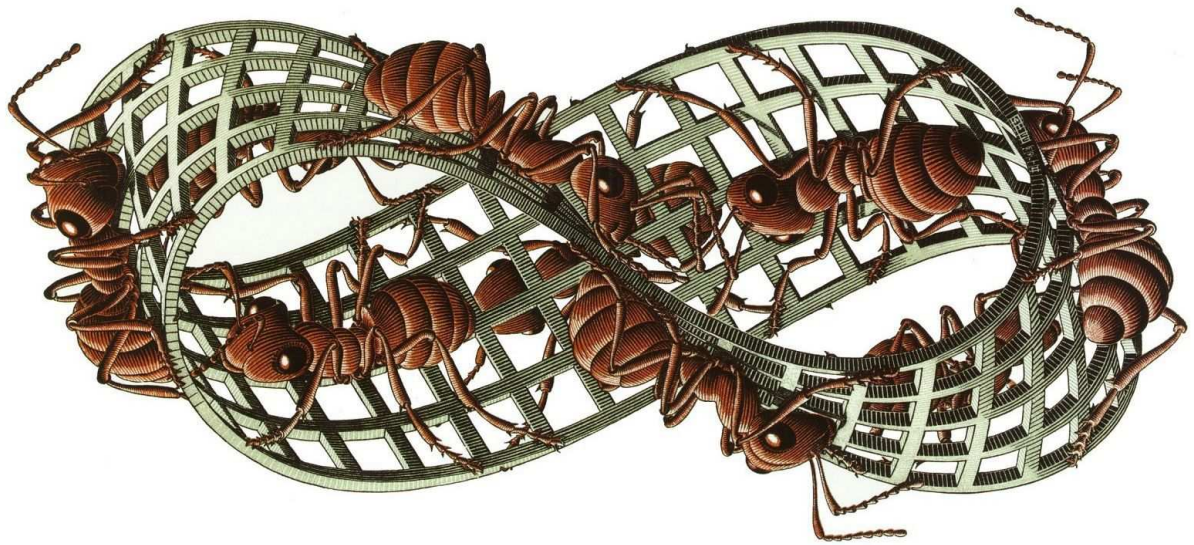
Hiermit erkläre ich an Eides statt, dass ich die vorliegende Arbeit selbständig und nur mit den angegebenen Hilfsmitteln angefertigt habe. Die Arbeit wurde von mir nicht an anderer Stelle eingereicht. Dies ist mein erster Promotionsversuch.

Kiel, 21.4.2008

---

Felix Köhler





M. C. Escher: Möbius Strip II

Said the ant to his friends: I declare!  
This is a most vexing affair.  
We've been ,round and round'  
But all we found  
Is the other side just isn't there.

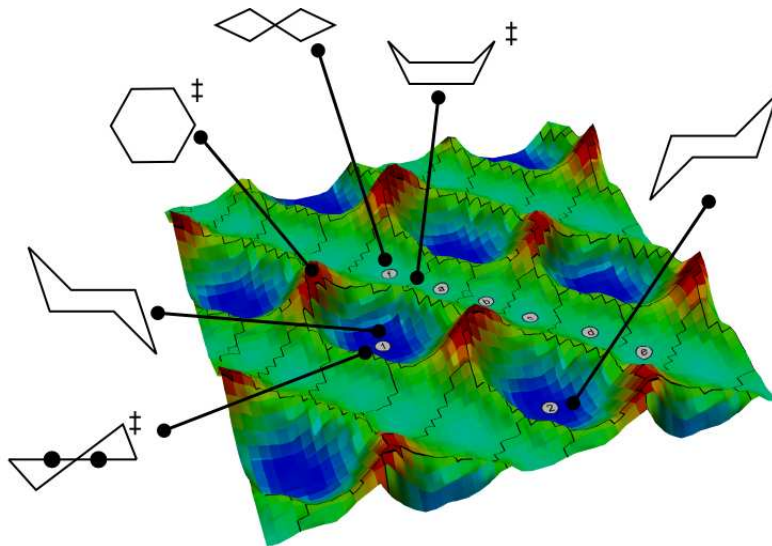
Die vorliegende Arbeit wurde von Dezember 1999 bis April 2008 am Institut für Organische Chemie der TU Braunschweig sowie am Institut für Organische Chemie der Universität zu Kiel unter der Anleitung von Herrn Prof. Dr. R. Herges angefertigt.

Ich möchte mich bei Prof. R. Herges ausdrücklich für die Möglichkeit bedanken, eine Arbeit anfertigen zu können die in ihrer stark theoretischen Ausrichtung deutlich außerhalb des Gebiets der klassischen organischen Chemie liegt und mir die Chance eröffnet hat, meinen Horizont als Naturwissenschaftler zu erweitern. Dank seiner ständigen Diskussionsbereitschaft, seiner Offenheit gegenüber Neuem sowie seiner großen menschlichen Loyalität gilt hoffentlich: „Was lange währt, wird endlich gut.“

Meiner Tante,  
Dr. Roswitha Neumann  
und meinen Eltern.

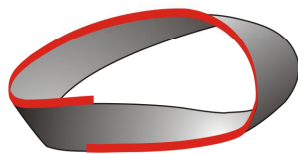
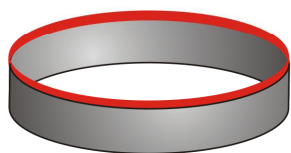
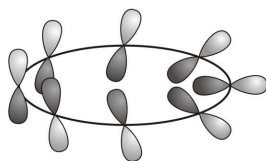
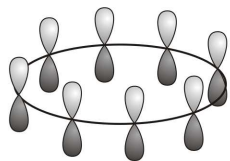
## Zusammenfassung der Doktorarbeit: Konformations- und Konfigurationshyperflächen - Generierung, Eigenschaften und Darstellung

Große flexible Moleküle, wie z.B. Proteine, Kohlenhydrate und andere Naturstoffe besitzen viele Freiheitsgrade und es ist in der Regel schwierig, die stabilsten und damit biochemisch wichtigen Konformationen und deren Umwandlungspfade zu finden. Ein sehr häufig verwendetes Verfahren um die Konformationsräume solcher molekularen Systeme theoretisch zu untersuchen ist die Moleküldynamik (MD). Nachteil dieses Verfahrens ist, dass eine grosse Menge Daten erzeugt wird, die hochredundant und multidimensional sind. In der vorliegenden Arbeit wird zum ersten Mal der Ansatz verfolgt, diese Daten mithilfe eines künstlichen, neuronalen Netzes vom Kohonen-Typ (welches der Hirnfunktion des visuellen Kortex ähnelt) zu untersuchen. Dabei werden zweidimensionale Karten der Konformationshyperfläche des Systems erhalten, die ähnlich wie topologische Landkarten, konformative Minima und deren Verbindungswege visuell anschaulich zeigen.



Neuronale Netzkarte des Cyclohexans als gut untersuchte Modells substanz

Die räumliche Struktur beeinflusst auch wichtige Moleküleigenschaften, wie z.B. die Aromatizität in cyclisch konjugierten Ringen (Annulenen). 1964 hatte der theoretische Chemiker E. Heilbronner vorhergesagt, dass Annulene, die in sich um  $180^\circ$  verdreht sind, d.h. eine Möbius-Topologie besitzen mit  $4n$  Elektronen aromatisch sind und damit gegen die Hückel- $(4n+2)$ -Regel verstoßen.



Hückel

Möbius

aromatisch mit  $4n+2$  Elektronen

aromatisch mit  $4n$  Elektronen

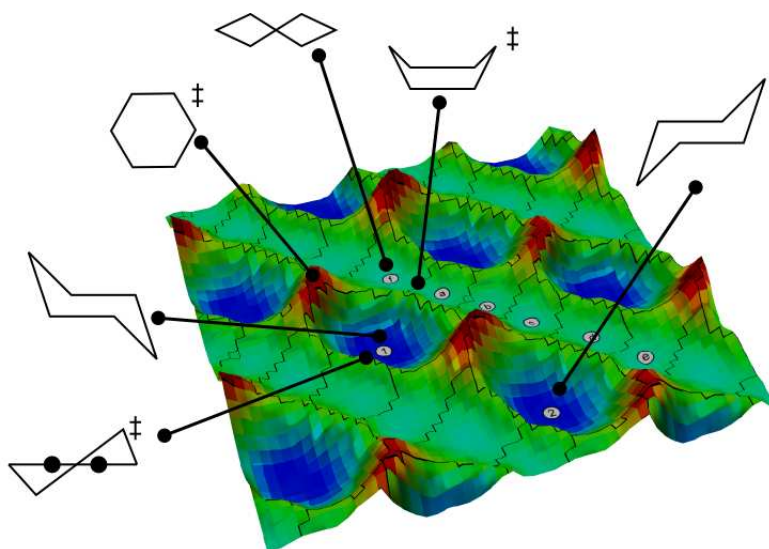
Trotz zahlreicher Versuche konnte bis vor kurzem kein experimenteller Nachweis für diese Hypothese gefunden werden. Größere Annulene-Ringe besitzen sehr viele mögliche Konformationen und Konfigurationen (z.B. [30]Annulene : 17920860). Um die Möbius-Nadeln im „Annulene-Heuhaufen“ zu finden wurde ein Programm entwickelt, welches automatisch Isomere der

Annulene generiert und auf Möbius-Topologie und Aromatizität untersucht. Unter einigen Annulenen wurden in der Tat stabile Möbius-Isomere gefunden, die möglicherweise interessante Synthesziele darstellen.



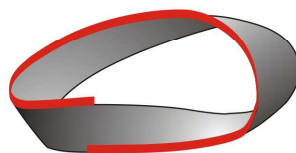
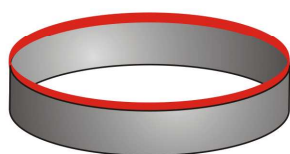
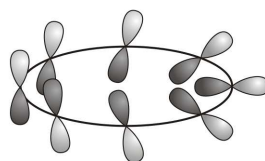
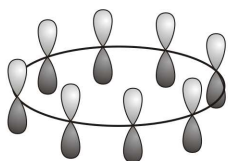
## Summary: Conformational and Configurational Hypersurfaces - Generation, Properties and Visualization

Large and flexible molecules, like proteins, carbohydrates and other natural products exhibit numerous degrees of freedom. It is generally quite difficult to find the most stable, and biochemically most active conformations and their paths of interconversion. Application of Molecular Dynamics (MD) has been established as one of the most important theoretical techniques to study the conformational space of these molecular systems. The sheer amount of highly redundant and multidimensional data produced by MD calculations constitutes a key problem of this method. This thesis will present the first application of artificial neural networks of the Kohonen type (which closely resemble the organization of the visual cortex) to investigate this wealth of data. Thus, two-dimensional representations of the conformational hypersurface are obtained, which, like topological maps, visualize conformational minima and interconversion pathways in an intuitive way.



Neural network map of cyclohexane as a well studied model substance

The three-dimensional molecular structure also influences important molecular properties, such as the aromaticity in cyclic, conjugated rings (annulenes). In 1964 the theoretical chemist E. Heilbronner suggested that annulenes, which are twisted in themselves by  $180^\circ$  in a Möbius topology, should be aromatic with  $4n$  electrons, thus, breaking the Hückel  $(4n+2)$ -rule.



Hückel

aromatic with  $4n+2$  electrons

Möbius

aromatic with  $4n$  electrons

Despite of numerous attempts, until very recently no experimental proof of this hypothesis was found. For large annulenes, a numerous conformations and configurations exist (e.g. [30]annulene). To find the "Möbius needles" in the "annulene haystack" an algorithm was developed and implemented in a program which automatically generates isomers and analyses them for Möbius topology and

aromaticity. Indeed, several stable Möbius isomers have been discovered which can be considered as interesting synthetic targets.



# Inhaltsverzeichnis

1	Einleitung und Aufgabenstellung .....	1
1.1	Die Konformation als Strukturmerkmal.....	1
1.2	Methoden der Konformationsuche.....	9
1.2.1	Systematische Suche im Konformationsraum .....	11
1.2.2	Moleküldynamiksimulationen.....	12
1.2.3	Monte-Carlo-Simulationen .....	14
1.3	Analyse hochdimensionaler Daten.....	15
1.3.1	Recurrence Analysis (RCA) .....	17
1.3.2	Multidimensional Scaling .....	18
1.3.3	Clustering Analysis (CA) .....	18
1.3.4	Principal Component Analysis (PCA) .....	19
1.3.5	Neuronale Netze.....	21
1.3.5.1	Neuronale Netze in Biologie und Informatik .....	22
1.3.5.2	Netzwerkarchitekturen .....	25
1.4	Aromatizität .....	26
1.4.1	Aromatizität als historischer Begriff.....	27
1.4.2	Hückel- und Möbiusaromatizität in Annulenen .....	29
1.4.3	Energiekriterien: ISE <sub>II</sub> , FISE .....	30
1.4.4	Magnetische Kriterien: NICS, MAGS, FISENICS, FISEMAGS..	36
1.4.5	Strukturelle Kriterien: HOMA, Julg-Index, GEO, EN .....	38
1.4.6	Anisotropie der induzierten Stromdichte (ACID).....	39
1.5	Aufgabenstellung .....	41
1.5.1	Teil 1: Auswertung konformativer Untersuchungen mit Hilfe von Kohonenkarten.....	41
1.5.2	Teil 2: Aromatizität von [4n]Annulen-Isomeren .....	42
2	Kohonnennetze zur Konformationsanalyse .....	44
2.1	Topologie der Kohonenkarte.....	44
2.2	Grafische Repräsentation der Kohonenkarte .....	46
2.3	Training von Kohonnennetzen – allgemeiner Teil .....	48
2.4	Training von Kohonnennetzen zur Konformationsanalyse.....	55
2.5	Qualitätskontrolle der Kohonenprojektion .....	58
2.5.1	RMSG .....	60
2.5.2	RMSDCF .....	60
2.5.3	RMSDSN.....	61
2.5.4	MTS.....	61
2.6	Grundfunktionalität von Kohonenkarten .....	61
2.7	Die NEURO-Programmsuite.....	65
2.7.1	Module als Organisationsprinzip .....	65
2.7.1.1	CONFSEARCHIX.....	66
2.7.1.2	PARAMETRIX .....	66
2.7.1.3	NORMALIX .....	67

2.7.1.4	KOHONIX.....	67
2.7.1.5	PROTOTYPIX.....	67
2.7.1.6	COLORIX.....	67
2.7.1.7	MINIX1.....	67
2.7.1.8	COLORIX1.....	67
2.7.1.9	MINIX2.....	68
2.7.1.10	COLORIX2.....	68
2.7.1.11	VISUALIX.....	68
2.7.1.12	SURFACEIX.....	68
2.7.1.13	EXTRACTIX.....	68
2.7.1.14	PROJECTIX.....	68
2.7.1.15	TRAJECTORIX.....	68
2.7.2	Übersicht der Kernprogramme:.....	69
2.8	Evaluierung des Netztrainings.....	71
2.8.1	n-Butan.....	72
2.8.1.1	Tutorial: MD-Rechnungen mit HYPERCHEM.....	73
2.8.1.2	Tutorial: Training einer Kohonenkarte.....	75
2.8.1.3	Tutorial: Training im Batchbetrieb.....	86
2.8.1.4	Tutorial: Hypersurface Explorer.....	88
2.8.1.5	Training der Grid-Hyperfläche.....	91
2.8.1.5.1	Variation der epochs.....	92
2.8.1.5.2	Variation von $f_0$ .....	95
2.8.1.5.3	Variation von $\Delta_{\max}(1)$ .....	98
2.8.1.5.4	Variation von $\eta(\xi)$ .....	99
2.8.1.5.5	Variation von NETX, NETY.....	101
2.8.1.5.6	Variation der Abstandsfunktion $\vartheta(\zeta)$ .....	102
2.8.1.5.7	Fazit.....	104
2.8.1.6	Beurteilung der Qualitätskriterien.....	105
2.8.1.7	MD Butan, Training der Diederwinkel.....	107
2.8.1.8	MD Butan, Training aller Distanzen.....	112
2.8.1.9	MD Butan, Training der Schweratomabstandsmatrix.....	115
2.8.2	n-Pentan.....	118
2.8.2.1	Training der Grid-Hyperfläche.....	119
2.8.2.1.1	Variation der epochs.....	120
2.8.2.1.2	Variation von $f_0$ .....	121
2.8.2.1.3	Variation von $\Delta_{\max}(1)$ .....	123
2.8.2.1.4	Variation von $\eta(\xi)$ .....	125
2.8.2.1.5	Variation von NETX, NETY.....	126
2.8.2.1.6	Variation der Abstandsfunktion $\vartheta(\zeta)$ .....	128
2.8.2.1.7	Fazit.....	129
2.8.2.2	MD Pentan, Training der Diederwinkel.....	130
2.8.2.3	MD Pentan, Training aller Distanzen.....	134

2.8.2.4	MD Pentan, Training der Schweratomdistanzen .....	137
2.8.3	n-Hexan.....	140
2.8.3.1	Training der Grid-Hyperfläche .....	143
2.8.3.1.1	Variation der epochs.....	144
2.8.3.1.2	Variation von $f_0$ .....	146
2.8.3.1.3	Variation von $\Delta_{\max}(1)$ .....	148
2.8.3.1.4	Variation von $\eta(\xi)$ .....	150
2.8.3.1.5	Variation von NETX, NETY .....	151
2.8.3.1.6	Variation der Abstandsfunktion $\vartheta(\zeta)$ .....	152
2.8.3.1.7	Fazit.....	156
2.8.3.2	MD Hexan, Training der Diederwinkel.....	157
2.8.3.3	MD Hexan, Training 30 zufälliger Distanzen .....	160
2.8.3.4	MD Hexan, Training der Schweratomdistanzen.....	163
2.8.4	n-Heptan.....	166
2.8.4.1	Training der Grid-Hyperfläche .....	167
2.8.4.1.1	Variation der epochs.....	168
2.8.4.1.2	Variation von $f_0$ .....	169
2.8.4.1.3	Variation von $\Delta_{\max}(1)$ .....	173
2.8.4.1.4	Variation von $\eta(\xi)$ .....	176
2.8.4.1.5	Variation von NETX, NETY .....	178
2.8.4.1.6	Variation der Abstandsfunktion $\vartheta(\zeta)$ .....	179
2.8.4.1.7	Fazit.....	182
2.8.4.2	MD Heptan, Training der Diederwinkel .....	183
2.8.4.3	MD Heptan, Training 42 zufälliger Distanzen.....	186
2.8.4.4	MD Heptan, Training der Schweratomdistanzen .....	189
2.8.5	n-Oktan .....	191
2.8.5.1	Training der Grid-Hyperfläche .....	193
2.8.5.1.1	Variation der epochs.....	193
2.8.5.1.2	Variation von $f_0$ .....	195
2.8.5.1.3	Variation von $\Delta_{\max}(1)$ .....	197
2.8.5.1.4	Variation von $\eta(\xi)$ .....	199
2.8.5.1.5	Variation von NETX, NETY .....	200
2.8.5.1.6	Variation der Abstandsfunktion $\vartheta(\zeta)$ .....	201
2.8.5.1.7	Fazit.....	205
2.8.5.2	MD Oktan, Training der Diederwinkel .....	205
2.8.5.3	MD Oktan, Training 51 zufälliger Distanzen.....	207
2.8.5.4	MD Oktan, Training der Schweratomdistanzen .....	210
2.8.6	Cyclohexan .....	212
2.8.6.1	Training der Grid-Hyperfläche .....	215
2.8.6.1.1	Variation der epochs.....	217
2.8.6.1.2	Variation von $f_0$ .....	219
2.8.6.1.3	Variation von $\Delta_{\max}(1)$ .....	221

2.8.6.1.4	Variation von $\eta(\xi)$ .....	223
2.8.6.1.5	Variation von NETX, NETY .....	224
2.8.6.1.6	Variation der Abstandsfunktion $\vartheta(\zeta)$ .....	225
2.8.6.1.7	Fazit.....	228
2.8.6.2	MD Cyclohexan, Training der Diederwinkel .....	229
2.8.6.3	MD Cyclohexan, Training 26 zufälliger Distanzen.....	232
2.8.6.4	MD Cyclohexan, Training der Schweratomdistanzen .....	234
2.8.6.5	Trajektorien Cyclohexan .....	236
2.8.7	Cis-Decalin .....	239
2.8.7.1	MD cis-Decalin, Training der Diederwinkel.....	239
2.8.7.2	MD cis-Decalin, Training 77 zufälliger Distanzen .....	243
2.8.7.3	MD cis-Decalin, Training der Schweratomdistanzen.....	246
2.8.8	Dibromcycloocten .....	249
2.8.8.1	Konformationsuche (HYPERCHEM conformational search) ...	249
2.8.8.2	MD 800 K.....	252
2.8.8.3	Konformationsräume im Vergleich.....	254
2.8.8.4	Topologiebeziehungen .....	256
2.8.9	Hexabenzocoronene (HBC) .....	257
2.8.9.1	Perbromo-HBC .....	257
2.8.9.2	Permethoxy-HBC .....	258
2.8.9.2.1	Konformationsanalyse (HYPERCHEM conformational search) .....	259
2.8.9.2.2	Moleküldynamik.....	260
2.8.9.2.3	Konformationsräume, TRAJECTORIX-Projektion .....	261
3	Aromatizität von [4n]Annulen-Isomeren.....	263
3.1	Einleitung .....	263
3.2	Die KONAN-Analyse.....	265
3.3	CANE-Deskriptoren .....	267
3.4	Übersicht der KONAN-Analyse.....	269
3.4.1	Konformationsuche.....	271
3.4.2	Verzeichnisstruktur .....	272
3.5	[12]Annulen .....	274
3.5.1	Konformationsuche.....	274
3.5.2	Umwandlung der Daten, Dublettenreduktion.....	274
3.5.3	PM3 single-point .....	277
3.5.4	DFT-Optimierung .....	278
3.5.5	NICS-Rechnungen .....	280
3.5.6	MAGS-Rechnungen .....	281
3.5.7	FISE-Rechnungen.....	283
3.5.8	FISE-NICS und -MAGS.....	284
3.5.9	Korrelationsgraphen.....	287
3.5.9.1	Strukturelle Kriterien.....	288

3.5.9.2	Energetische Kriterien .....	289
3.5.9.3	Magnetische Kriterien.....	290
3.5.10	AICD-Plots .....	294
3.6	[16]Annulen.....	299
3.6.1	MM+ Startstrukturen .....	300
3.6.1.1	Konformationssuche .....	300
3.6.1.2	PM3 und Optimierung (B3LYP/6-31G*) .....	300
3.6.1.3	NICS/MAGS/FISE-Rechnungen.....	302
3.6.1.4	Strukturelle Kriterien .....	304
3.6.1.5	Energetische Kriterien .....	305
3.6.1.6	Magnetische Kriterien.....	307
3.6.1.7	Principal Component Analysis (PCA) .....	311
3.6.1.8	Vergleich mit Literaturdaten .....	316
3.6.2	Andere Kraftfelder .....	319
3.7	Das erste Möbiusannulen.....	320
4	Zusammenfassung und Ausblick.....	325
4.1	Kohonennetze zur Konformationsanalyse (NEURO).....	325
4.1.1	Standardtrainingsparamter.....	325
4.1.2	Vergleich mit konkurrierenden Verfahren.....	326
4.1.2.1	Stärken .....	326
4.1.2.2	Schwächen .....	327
4.1.3	Ausblick .....	328
4.2	Aromatizität von Annulen-Isomeren.....	329
4.2.1	Zusammenfassung .....	329
4.2.2	Ausblick .....	331
5	Appendix A: Verwendete Programme und Hilfsmittel.....	332
5.1	Externe Hilfsmittel .....	332
5.1.1	Modellingprogramme.....	332
5.1.2	Compiler und Interpreter.....	333
5.1.3	Externe Code-Bibliotheken (libraries) .....	333
5.1.4	eigene Code Bibliotheken.....	333
5.1.4.1	HYPERSNAP.....	333
5.1.4.2	Net .....	334
5.2	Hilfsskripte für HYPERCHEM .....	334
5.2.1	FIXROT.TK .....	335
5.2.2	Grid.tcl .....	336
5.2.3	HCL.TK .....	336
5.2.4	LINTORS.TK (LINEAR TORSIONS) .....	336
5.2.5	MAKE_DISTANCE_MATRIX.TK .....	336
5.2.6	PRINT_SELECTION_VALUES_GUI.TK.....	337
5.2.7	RELOAD.TK .....	337
5.2.8	RINGTORS.TK (RING TORSIONS) .....	337

5.2.9	SCRIPTS.TK .....	337
5.2.10	SCRIPT_LOADER.TK .....	337
5.2.11	SYSTEMATIC_SEARCH_PERBROMO_HBC.TK.....	337
5.2.12	TCL.TK.....	337
5.2.13	XYZ2COM.TCL.....	338
5.3	Eigene Programme.....	338
5.3.1	ANALYSE.EXE [NEURO-SUITE] .....	338
5.3.2	ASSIGN.EXE [NEURO-SUITE] .....	339
5.3.3	CLASSIFY_BUTAN_SNP.EXE.....	339
5.3.4	CLASSIFY_PENTAN_SNP.EXE.....	340
5.3.5	CLASSIFY_SNP_CONFORMATIONS.EXE.....	340
5.3.6	CYCHEX_TORS.EXE .....	340
5.3.7	EXAKTE_LOESUNG.EXE .....	340
5.3.8	FIX_ROTATION.EXE.....	340
5.3.9	GENERATE_SNAP_HEADERS .....	341
5.3.10	HCS2SNP.EXE .....	341
5.3.11	JSNAPS.EXE.....	341
5.3.12	MAKE_2ND [NEURO-SUITE].....	341
5.3.13	MINIMIZE.EXE [NEURO-SUITE].....	342
5.3.14	NEURO.EXE [NEURO-SUITE] .....	342
5.3.15	NN.EXE [NEURO-SUITE].....	343
5.3.16	POKESNAP.EXE .....	343
5.3.17	SEQLOG2XYZ.EXE.....	344
5.3.18	SNP2XYZ.EXE .....	344
5.3.19	SUBSNAP.EXE .....	344
5.3.20	TRANSLATE_SNAP.EXE.....	345
5.3.21	XYZ2CANE.EXE .....	345
5.3.22	XYZ2SNP.EXE .....	346
6	Appendix B: Distribution und supplementary material.....	347
7	Appendix C: Anmerkungen und Literaturverzeichnis.....	349







# 1 Einleitung und Aufgabenstellung

Diese Arbeit beschäftigt sich mit zwei großen Themengebieten unter dem Oberbegriff der Konformationsanalyse. Das erste umfasst die Generierung des Konformationsraums einfacher und komplexer Moleküle mittels Molekulardynamikrechnungen und anderer Methoden sowie der anschließenden Darstellung der Ergebnisse im zweidimensionalen Raum mit Hilfe von neuronalen Kohonennetzen. Die dafür verwendeten Methoden und Programme wurden im Rahmen dieser Arbeit entwickelt. Diese Programme, die verwendeten Algorithmen, ihre Anbindung an bestehende chemische Rechenprogramme, Anwendung auf ausgesuchte Beispiele sowie die Limitierung der Methode sind Gegenstand des ersten Hauptteils dieser Arbeit.

Im zweiten Teil wird die Untersuchung von zyklischen Annulenen auf Aromatizität beschrieben. Insbesondere die Einordnung von geladenen und ungeladenen Annulenen mit  $4n$   $\pi$ -Elektronen in Möbiuskonfiguration wird durchaus noch kontrovers diskutiert.<sup>[1]</sup> Dabei werden im Rahmen dieser Arbeit ebenfalls neue Methoden zur effizienten Erzeugung und Auswertung einer großen Anzahl von Konformeren sowie Konfigurationsisomeren mit Hilfe des Programms Gaussian03<sup>[2]</sup> entwickelt. Diese Arbeit soll zur Klärung der Frage beitragen, ob Möbiusaromatizität in diesen Molekülen tatsächlich vorliegt.

## 1.1 Die Konformation als Strukturmerkmal

Konformationen sind Isomere eines Moleküls, die sich ohne das Brechen oder Neuformieren einer Bindung ineinander überführen lassen. Die Grenze zur Konfigurationsisomerie wie z.B. *cis-trans* Doppelbindungsisomerie verläuft fließend. Normalerweise wechseln Moleküle die Konformation jedoch durch das Rotieren von Diederwinkeln um Einfachbindungen, was üblicherweise mit sehr geringen Aktivierungsbarrieren abläuft. In kleineren Molekülen laufen Konformationsänderungen bei Raumtemperatur typischerweise so schnell ab, dass Konformere nicht isoliert werden können. Erst in großen Molekülen mit vielen relativ starren Einfachbindungen zeigen sich Konformationen als persistent.

Konformationen als prägende Strukturelemente von Molekülen sind erst in neuerer Zeit bei Untersuchungen der großen Biomoleküle in den Fokus der wissenschaftlichen Diskussion geraten. Diese Biomoleküle wie Desoxyribonukleinsäure (DNA) und Proteine sind im Gegensatz zu vielen künstlichen

Polymeren mit vielen starren Einfachbindungen ausgestattet. In Proteinen sind dafür zum großen Teil die Amidbindungen mit ihrem partiellen Doppelbindungscharakter verantwortlich. Um die Einordnung der Konformationsisomerie in das chemische Weltbild der Struktur der Materie zu verstehen, ist ein kurzer historischer Abriss der Entwicklung der Chemie und insbesondere ihrer Gedanken- und Erklärungsmodelle hilfreich.



**Abbildung 1: Die Irrtümer der Alchemisten. Holzschnitt aus Sebastian Brants „Narrenschiff“ (Basel 1494).<sup>[3]</sup>**

Der Aufstieg der mittelalterlichen Alchemie zur Naturwissenschaft Chemie präsentiert sich dabei in mehrerlei Hinsicht als faszinierende Entwicklung. Am Anfang beeindrucken hauptsächlich Einsatz und Fleiß, mit denen weitgehend ohne theoretischen Hintergrund in mühevollen Versuchen wichtige Entdeckungen wie das Schießpulver und Porzellan gemacht wurden. Später gewannen theoretische Vorhersagen und ihre gezielte Verifizierung bzw. Falsifi-

zierung an Einfluss, auch wenn aktuell die Empirik in Form der kombinatorischen Chemie an Bedeutung zurückgewonnen hat.

Als Übergang zur modernen Naturwissenschaft kann die Postulierung des Daltonschen Atommodells 1808 gesehen werden<sup>[4]</sup> mit der Erforschung der chemischen Bindung als logischer Konsequenz. Wenn man diese Geschichte aus dem Blickwinkel der verwendeten Struktur- und Bindungsmodelle betrachtet, wird dabei ein Leitmotiv sichtbar, das mit der Komplexität des chemischen Verständnisses zusammenhängt. Die Repräsentation der verwendeten Daten nimmt dabei stetig an Dimensionalität zu.

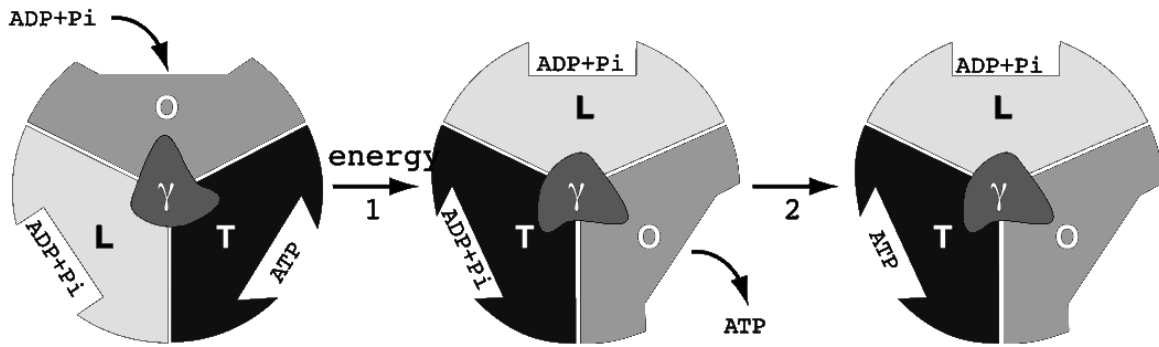
Einige der wesentlichen Meilensteine der Chemie und ihre Auswirkungen auf das chemische Weltbild sind in Tabelle 1 zusammengefasst.

**Tabelle 1: Meilensteine der Chemie**

1794	Joseph-Louis Proust stellt sein nach ihm benanntes Gesetz der konstanten Proportionen vor, welches eine wesentliche Grundlage für die Entwicklung des Daltonschen Atombegriffs darstellt.	Chemische Summenformeln, <b>1D</b>
1808	Dalton stellt sein Atommodell vor und stellt damit die theoretische Basis für kommende analytische Verfahren bereit. <sup>[4]</sup>	
1818	Jöns Jakob Berzelius publiziert eine Liste der Atomgewichte und führt die heute gebräuchliche Symbolschreibweise für Summenformeln in der Chemie ein.	
1831	Justus von Liebig gelingt die Entwicklung des 5-Kugelapparates, der die Elementaranalyse auf eine zuverlässige Basis stellt und die Ermittlung der Summenformel organischer Substanzen zur Standardmethode macht.	
1858	Friedrich August Kekulé postuliert die Vierwertigkeit des Kohlenstoffs und das Vorhandensein von Kohlenstoff-Kohlenstoff-Bindungen. Dies kann als Basis für die Darstellung von Molekülen in zweidimensionaler Form angesehen werden.	Chemische Konstitution, <b>2D</b>
1864	Kekulé erklärt, anekdotisch angeregt durch den Traum einer sich selbst in den Schwanz beißenden Schlange, die Struktur des Benzols als einen aus sechs Kohlenstoff-Atomen bestehenden symmetrischen Ring. Dies ist der Durchbruch der zweidimensionalen Darstellung der Moleküle als vorherrschendes Modell.	

1859	Louis Pasteur gelingt die erste Racematspaltung durch Trennen von Kristallen der Weinsäure. Damit existiert ein Beweis für die Existenz von Molekülen als „Bild und Spiegelbild“.	Molekülstruktur, chemische Konfiguration, Konformationen, <b>3D</b>
1875	Jacobus Henricus van 't Hoff erklärt in seinem Werk <i>La chimie dans l'espace</i> die optische Aktivität durch das Tetraedermodell des Kohlenstoff. Dies ist der Anfang der 3D-Strukturdarstellung	
1914	Max von Laue bekommt den Nobelpreis der Physik für die Entdeckung der Röntgenbeugung an Kristallen. Damit können dreidimensionale Kristallstrukturen durch Röntgenstrukturanalyse ermittelt werden. Dafür gibt es einen Nobelpreis nur ein Jahr später für Sir William Henry Bragg und William Lawrence Bragg.	
1931	Ernst Hückel habilitiert mit bahnbrechenden Arbeiten über das Benzol und entwickelt dabei die Hückel Näherung und die bekannte $4n+2$ Hückelregel für aromatische Ringsysteme.	
1953	Der Nobelpreis für Medizin 1962 geht an Francis Harry Compton Crick, James Dewey Watson und Maurice Hugh Frederick Wilkins für die Aufklärung der dreidimensionalen Struktur der Desoxyribonukleinsäure (DNA). Der wesentliche experimentelle Hinweis war dabei nach einem der Beteiligten <sup>[5]</sup> eine Röntgenbeugungsphotographie von Rosalind Franklin. Dies kann als der Beginn des „Röntgenstrukturzeitalters“ in der Biochemie angesehen werden. Gerade im Bereich Biochemie richtet sich der Fokus auf die Struktur der Moleküle im Raum als entscheidendes bisher missachtetes Strukturelement, welches viele Moleküleigenschaften bestimmt, wobei die Grenze zwischen Chemie und makroskopischer Welt erreicht wird. Die Röntgenstrukturanalyse entwickelt sich zur „harten“ Strukturbeweismethode der Wahl.	

ca. 1950 – 1960	<p>Arbeiten über die Renaturierung der <i>bovine pancreatic ribonuclease</i> (PDB 2AAS)<sup>[6]</sup> führen Anfinsen zum Postulat,<sup>[7]</sup> dass die dreidimensionale Struktur eines natürlichen Proteins in seinem normalen physiologischen Milieu (Lösemittel, pH-Wert, Ionenstärke, Anwesenheit anderer Substanzen wie Koenzyme, Temperatur etc.) derjenigen Konformation mit der niedrigsten Gibbs-Energie für das gesamte System ist.</p> <p>Dies wird zum wichtigen Paradigma bei theoretischen Untersuchungen und Versuchen, die Faltung eines Proteins aus seiner Sequenz vorherzusagen.<sup>[8]</sup></p>	Konformerensemble, Moleküldynamiktrajektorien, ND
1958 – 1960	<p>Erste Moleküldynamikrechnungen im Bereich der theoretischen Physik an Edelgasen, zunächst ohne Modellierung von Bindungen<sup>[9],[10],[11]</sup> eröffnen die Möglichkeit, Moleküleigenschaften aus sich bewegenden Molekülen zu ermitteln. Dies führt zur Bestimmung von bisher per Modell nicht erfassbaren thermodynamischen Größen.</p>	
1997	<p>Der Nobelpreis für Chemie 1997 geht an Paul D. Boyer, John E. Walker und Jens C. Skou für die Entdeckung der ATPase und der Aufklärung der Funktionsweise der Adenosintriphosphat (ATP) - Synthese aus Adenosindiphosphat (ADP) und anorganischem Phosphor (P<sub>i</sub>). Wie in Abbildung 2<sup>[12]</sup> ersichtlich, enthält das Protein sequenzidentische <math>\beta</math>-Untereinheiten, die sich nur in ihrer Konformation unterscheiden. Mittels eines Protonengradienten wird dem System Energie bei Durchschleusen von 3, evtl. 4 Protonen Energie zugeführt und die F1-ATPase zur Rotation gebracht.<sup>[13],[14]</sup> Dies zwingt die Untereinheiten zu Konformationsänderungen, im Verlauf derer die endergonische Reaktion des gebundenen ADP zu ATP nicht nur katalysiert, sondern außerdem energetisch unterstützt wird.</p>	



**Abbildung 2:** Schematischer Ablauf der Synthese von ATP aus ADP und Phosphat (Pi) im Protein F1-ATPase der Mitochondrienmembran<sup>[12]</sup>

Der Übergang vom Einfachen zum Komplexen beim Strukturbegriff der Chemie lässt sich mit Hinblick auf die Dimensionalität des verwendeten Modells damit wie folgt zusammenfassen: Die Identifizierung der Elemente und die Entwicklung der Elementaranalyse erlaubte eine Charakterisierung der Substanzen durch Angabe der Summenformel sowie physikalischer Messgrößen wie Schmelzpunkt und Siedepunkt (eindimensional 1D). Die Entdeckung von Isomeren erfordert ein Modell der chemischen Bindung. Die Ermittlung der Bindungsverhältnisse (Konnektivitäten) ergibt die Struktur im zweidimensionalen Raum (2D), welche für viele kleine Moleküle ein ausreichendes und erfolgreiches Modell darstellt. Zusätzliche dreidimensionale Merkmale wie die Konfiguration chiraler Moleküle werden erkannt und in die zweidimensionale Abbildung eingearbeitet.

Bei den großen Biomolekülen wird spätestens mit Entwicklung der Röntgenstrukturanalyse klar, dass die räumliche Struktur und die Konformation eines Moleküls auch einen signifikanten Einfluss auf sein physiologisches Verhalten und seine Reaktivität haben können. Sowohl die Analysemethoden (NMR, ESR, X-RAY) als auch die theoretischen Behandlungsmethoden molekularer Systeme werden immer ausgefeilter und etablieren den dreidimensionalen Strukturbegriff. Das Vorhandensein von immer mehr Rechenleistung ermöglicht im Weiteren die theoretische dynamische Betrachtung von Systemen. Forschungsergebnisse in der Biochemie ergeben die Notwendigkeit zur Betrachtung nicht nur der Minimumskonformation, sondern eines umfangreicheren Konformerensemble in Form einer hochdimensionalen Hyperfläche. Dies ist in der folgenden Tabelle 2 zusammen mit einigen der wichtigsten verwendeten informationstechnischen Hilfsmittel aufgetragen.



**Tabelle 2: Strukturmodelle in der Chemie**

ND	Strukturmodell	Anwendung	Datenbanken (Beispiele)
<b>1D</b>	Sequenz, Summenformel	eindeutige Identifikation von Substanzen unter Zuhilfenahme mehrerer „eindimensionaler“ Substanzmerkmale wie Summenformel oder Schmelzpunkt	Gmelin, Beilstein
<b>2D</b>	Konnektivität, Konfiguration	Struktur/Substruktursuche, Reaktionsplanung	Beilstein, Crossfire
<b>3D</b>	Kartesische Koordinaten	3D-Ähnlichkeitssuche ( <i>3D-similarity-search</i> ), Leitstruktursuche	Protein Data Bank (RCSB, früher PDB), <sup>[15]</sup> Cambridge Structural Database (CSD) <sup>[16]</sup>
<b>&gt;3D ND</b>	Mehrdimensionale Hyperflächen, Konformationsflexibilität, dynamische Systeme	Induced Fit, Drug Design, Docking Studien	NCI-DIS

Statische 3D-Strukturen beschreiben kleine organische und anorganische Strukturen sowie Reaktionsverläufe oft ausreichend genau. Klassische quantenmechanische Energieoberflächen werden durch stationäre Punkte, also statische dreidimensionale Strukturen, meist klar definiert (Ausnahmen sind extrem flache Hyperflächen, bei denen die Born-Oppenheimer-Näherung nicht mehr gilt).

Große Systeme wie Proteine zeichnen sich jedoch durch konformative Hyperflächen aus, welche zu komplex sind, um sie übersichtlich und aussagekräftig mit einem Satz an statischen Strukturen zu beschreiben. Bei der höherdimensionalen Darstellung von dynamischen komplexen Systemen sind dabei noch einige Hindernisse zu überwinden. Dies gilt sowohl für die rechnerische Behandlung als auch für die Darstellung und Verwendung der Ergebnisse von Analysen. Diese Probleme sind zum Teil so fundamental, dass eine vollständige Lösung auf absehbare Zeit nicht in Aussicht steht.

Für große Moleküle wie Proteine steigt der Rechenaufwand der theoretischen Untersuchung der Energiehyperfläche mit wachsender Anzahl an Atomen stark an. Schon bei relativ kleinen Proteinen (20-100 Aminosäuren (AS)) gibt es eine unüberschaubar große Anzahl von Energieminima. Aus informationstheoretischer Sicht gehört die Konformationsuntersuchung fernerhin zur Klasse der sogenannten np-vollständigen Probleme.<sup>[17],[18]</sup> Die np-Vollständigkeit (*np-completeness*,  $np=non\ polynomial$ ) eines Problems lässt sich formal beweisen. Ein np-vollständiges Problem kann nach heutigem Wissen nicht in Polynomialzeit gelöst werden. Die Zeit, die zur vollständigen Lösung benötigt wird, steigt exponentiell mit wachsender Größe eines skalierenden Faktors  $n$ , z.B. der Anzahl der Atome in einem Molekül, an. Dies führt zu einer Rechenzeitexplosion, die mit heutigen Maschinenarchitekturen nicht gelöst werden kann. Neuartige Rechnerarchitekturen mit dem Potential np-vollständige Probleme zu lösen werden zurzeit diskutiert in Form von DNA-Computern<sup>[19]</sup> und Quantencomputern.<sup>[20]</sup> Diese sind für alle praktischen Anwendungen allerdings den konventionellen Maschinen unterlegen. Algorithmen mit polynomialem Laufzeitverhalten verhalten sich bei Skalierung deutlich gutmütiger, auch wenn in der Praxis meist nur lineare, quadratische und kubische Algorithmen sinnvoll skalieren. Die Frage, ob die Klasse der np-vollständigen Probleme auf polynomiales Laufzeitverhalten reduziert werden kann, kurz  $np=p$ , ist außerdem immer noch ein ungelöstes Problem. Sollte hier ein Durchbruch geschaffen werden, hätte dies unmittelbare Auswirkungen auf die Berechenbarkeit aller np-vollständigen Probleme. Allerdings wird eine vollständige Lösung meist nicht benötigt. Viele Probleme lassen sich „ausreichend gut“ in akzeptabler Zeit behandeln, auch wenn die Ergebnisse nicht notwendigerweise ein Optimum darstellen.

Dies bedeutet, dass eine vollständige Untersuchung des Konformationsraums bereits bei relativ kleinen Oligopeptiden und pharmazeutisch wirksamen Molekülen zumindest mit dem derzeitigen Prinzip der Computer (von Neumann - Maschinen) und Stand der theoretischen Informatik nicht möglich ist. Durch Wahl passender Methoden und Parameter zur Einschränkung des zu untersuchenden Konformationsraums können unter Umständen aber Ergebnisse erhalten werden, die dem Anspruch auf für das Problem ausreichende Vollständigkeit genügen.

Ein weiteres gravierendes Problem besteht in der Komplexität der erhaltenen Ergebnisse. Daten aus Konformationsuntersuchungen zeichnen sich durch Redundanz und hohe Dimensionalität aus und sind damit nur äußerst schwer zugänglich für die menschliche Auffassungsgabe. Es fehlen Methoden zur

Visualisierung oder automatisierte Verfahren zur Aufarbeitung hochdimensionaler Daten, beides Themen dieser Arbeit.

## 1.2 Methoden der Konformationsuche

Im vorausgehenden Kapitel wurde die Notwendigkeit beschrieben, Flexibilität im Konformationsraum von Molekülen bei der Modellierung zu berücksichtigen. Der klassische Ansatz im *Molecular Modelling* ist allerdings statisch. Quantenmechanische Rechnungen ergeben lokale Lösungen der Schrödinger-Gleichung auf der Potentialhyperfläche, so genannte stationäre Punkte. Wenn sich Moleküleigenschaften aus Beiträgen mehrerer stationärer Punkte ergeben, so werden Gleichgewichtsverteilungen eines Satzes ausgewählter Strukturen zur Ermittlung herangezogen. Molekülmechanikrechnungen ergeben oft einen Satz an Minimumskonformationen, im kartesischen Koordinatensystem dargestellt durch einen Satz Koordinaten für jedes Atom des Systems (siehe (1.1)).

$$(1.1) \quad \mathbf{q} = (x_1, y_1, z_1, x_2, y_2, z_2, \dots)$$

Mittels einer Boltzmannstatistik lassen sich auch hier thermodynamische Eigenschaften berechnen. Allerdings ist dies für große Systeme mit vielen Freiheitsgraden nicht notwendigerweise ein korrekter Ansatz. Die Reduktion des Konformationsraums auf die Minima ist nur so lange gültig, wie man es mit einer „wohldefinierten“ Hyperfläche mit „tiefen“ und „engen“ Minima zu tun hat. Üblicherweise ist dies in komplexen Systemen mit vielen Freiheitsgraden nicht der Fall. Große Makromoleküle wie z.B. Polyethylenglykol oder große Alkane haben sehr viele sehr „weiche“ Torsionsschwingungen, die ohne größere Übergangsbarriere Minima verbinden, was eine ausgesprochen flache komplexe Hyperfläche zur Folge hat.

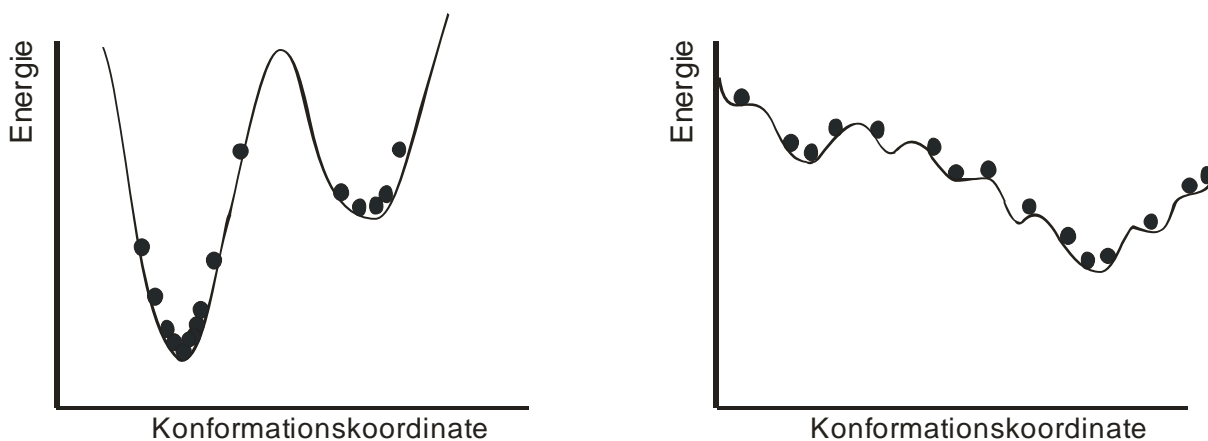


Abbildung 3: einfache „tiefe“ (links) und komplexe „flache“ Hyperfläche (rechts)

Während die einfache Hyperfläche als statistisches Mittel durch ihre stationären Punkte gut beschrieben wird, herrschen in der komplexen Hyperfläche Konformationsbereiche statt klar umrissener Punkte vor.

Außerdem gibt es grundsätzliche Probleme bei der Bestimmung von Minima in flachen Hyperflächen. In der Quantenmechanik bricht die Born-Oppenheimer-Näherung zusammen, die die Bewegung der Elektronen von der Bewegung der Kerne entkoppelt. Molekülmechanikrechnungen leiden darunter, dass von einer Ausgangskonfiguration oft je nach Optimieralgorithmus und anderen teilweise nicht deterministischen Effekten in der Berechnung unterschiedliche Minima gefunden werden. Sinnvoller ist es, ebenfalls die Momente der einzelnen Atome, wiederum ausgedrückt in kartesischen Koordinaten, zu betrachten:

$$(1.2) \quad \mathbf{p} = (p_{x,1}, p_{y,1}, p_{z,1}, p_{x,2}, p_{y,2}, p_{z,2}, \dots)$$

Damit ergibt sich dann der komplette Phasenraum zu

$$(1.3) \quad \mathbf{X}' = (\mathbf{q}, \mathbf{p}) = (x_1, p_{x,1}, y_1, p_{y,1}, z_1, p_{z,1}, x_2, p_{x,2}, y_2, p_{y,2}, z_2, p_{z,2}, \dots)$$

Die Wahrscheinlichkeit eines Systems, sich in einem bestimmten Punkt des Phasenraums aufzuhalten, ist bekannt aus der statistischen Thermodynamik:

$$(1.4) \quad P(\mathbf{q}, \mathbf{p}) = Q^{-1} e^{-E(\mathbf{q}, \mathbf{p})/k_B T}$$

mit der totalen Energie  $E$ , der Boltzmann-Konstante  $k_B$  und der Zustandssumme (englisch: *system partition function*)  $Q$ , wobei  $Q$  durch Gleichung (1.5) definiert ist.

$$(1.5) \quad Q = \int \int e^{-E(\mathbf{q}, \mathbf{p})/k_B T} d\mathbf{q} d\mathbf{p}$$

Die einfachste Methode der Betrachtung von Konformerensembles besteht darin, ein Gitter über den Phasenraum zu legen und für alle denkbaren Kombinationen die Energie auszurechnen. Bei konstanter Temperatur, Volumen und Zahl der Teilchen ergibt sich der Erwartungswert thermodynamischer Größen des Systems aus:

$$(1.6) \quad \langle A \rangle = \int \int A(\mathbf{q}, \mathbf{p}) P(\mathbf{q}, \mathbf{p}) d\mathbf{q} d\mathbf{p}$$

Die Gleichung (1.5) lässt sich jedoch nicht einfach analytisch lösen. Außerdem kommt es, wie von C. J. Cramer ausgeführt,<sup>[21]</sup> zur sog. kombinatorischen Explosion. Eine vollständige Betrachtung (*complete sampling*) des Phasenraums ist in größeren Systemen nicht einmal annähernd möglich. Je nach Art des zu

untersuchenden Systems ist dies aber auch nicht erforderlich. Der größte Bereich des Phasenraums ist unter normalen Bedingungen aufgrund der hohen Energie und des resultierenden minimalen Einflusses auf die Gesamtfunktion in (1.5) nicht populiert.

Es sind im Wesentlichen drei Methoden bekannt, um die wichtigen Bereiche des Phasenraums zu untersuchen: systematische Suche, Moleküldynamik und Monte-Carlo (wichtig bedeutet in diesem Zusammenhang „für die Eigenschaften des Systems repräsentativ“).<sup>[22]</sup> Je nach System, Komplexität des Problems, verwendeter Rechenmethode (Molekülmechanik/Quantenmechanik) und zusätzlich unterstützenden vorhandenen experimentellen Daten (hauptsächlich *NMR-restraints*) gibt es weitere Adaptionen der oben beschriebenen Verfahren. Als Beispiel seien hier die *umbrella-sampling technique*<sup>[23]</sup> zur Untersuchung von Systemen entlang vordefinierter Reaktionskoordinaten sowie der SHAKE<sup>[24]</sup> Algorithmus zum Einschränken des Konformationsraums in Hinblick auf per NMR Messung gewonnene *restraints* genannt. Eine ausführlichere Beschreibung findet sich in Teil 1.2.2, S. 12.

### 1.2.1 Systematische Suche im Konformationsraum

Die systematische Suche im Konformationsraum stellt historisch gesehen in Form der *grid-search* Suche die erste Methode dar, den Konformationsraum zu untersuchen. Von Vorteil ist es dabei, nicht den kompletten Phasenraum zu betrachten, sondern das System auf einen ausgesuchten Parameterraum zu reduzieren. Besonders Diederwinkel werden gerne herangezogen, da sie die größten Veränderungen beim Konformationswechsel zeigen.<sup>[25]</sup> Für wenig komplexe Systeme ist die *grid-search* Methode leistungsstark, da sie eine relativ vollständige Betrachtung liefert. Eine einfache Anwendung ist z.B. die Rotation des Butans um seine zentrale Bindung.

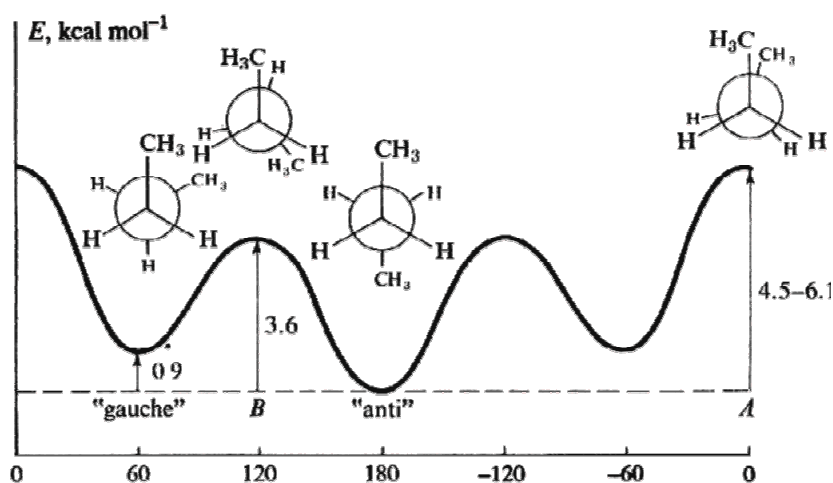


Abbildung 4: Potentielle Energie von Butan abhängig vom Diederwinkel<sup>[26]</sup>

Bei komplexeren Molekülen stößt diese Methode jedoch schnell an ihre Grenzen. Moderne systematische Suchalgorithmen arbeiten nicht mehr mit *grid-search* sondern ändern gezielt einzelne interne Koordinaten. Am Beispiel des Cycloheptadekans untersuchen Saunders, Guida *et al.*<sup>[27]</sup> die Leistungsfähigkeit verschiedener Methoden. Darunter sind verschiedene systematische Suchen sowie einige Moleküldynamik- (MD) und Monte-Carlo-Methoden (MC) im Koordinatenraum der Torsionswinkel der CC-Einfachbindungen. Schon bei diesem relativ überschaubaren Molekül kommt ein vollständiges Suchgitter aufgrund der „kombinatorischen Explosion“ nicht mehr in Frage. Als Ergebnis erhält man, dass die systematische Suche im Parameterraum, vor allem für Torsionswinkel, eine um Größenordnungen bessere, d.h. vollständigere und schnellere Abdeckung des gesamten Konformationsraums liefert und für kleinere Systeme die Methode der Wahl darstellt. Dies ist der Grund, warum sie im Rahmen von quantenmechanischen Rechnungen im zweiten Hauptteil dieser Arbeit anstelle von anderen Methoden verwendet wird.

### 1.2.2 Moleküldynamiksimulationen

Bei der MD-Simulation wird eine Trajektorie des Molekülsystems (Konfigurationen als Funktion der Zeit) durch die gleichzeitige Integration der Newtonschen Bewegungsgleichungen aller Atome des Systems iterativ erzeugt.<sup>[28]</sup>

$$(1.7) \quad \frac{d^2 \mathbf{r}_i(t)}{dt^2} = m_i^{-1} \cdot \mathbf{F}_i$$

$$(1.8) \quad \mathbf{F}_i = - \frac{\partial V(\mathbf{r}_1, \dots, \mathbf{r}_N)}{\partial \mathbf{r}_i}$$

Die auf ein Atom  $i$  wirkende Kraft wird mit  $\mathbf{F}_i$  und die Zeit mit  $t$  bezeichnet. MD-Simulationen erfordern die Berechnung des Gradienten der potentiellen Energie  $V(\mathbf{r})$ , die daher differenzierbar sein muss. Ansonsten können hier beliebige Ansätze gewählt werden. Üblich sind aufgrund des hohen Rechenaufwandes molekülmechanische Rechnungen, aber auch quantenmechanische Rechnungen wären im Prinzip möglich.

Kraftfelder definieren eine Reihe von Energiekomponenten, deren Differentiale dann Kräfte ergeben, die Termen der klassischen Mechanik sehr ähneln. Daher auch der alternative Name Molekülmechanik:

$$(1.9) \quad E_i = \sum E_{stretch} + E_{bend} + E_{tors} + E_{static} + E_{dipol} + E_{vdw} + \dots + E_{restraint}$$

Kraftfelder unterscheiden sich in Art und Anzahl der verwendeten Potentiale und/oder Parametern und Atomtypen.

Eine Übersicht über die in der Organischen Chemie (OC) und Biochemie (BC) wichtigsten Kraftfelder und Programme ist in der folgenden Tabelle zusammengestellt:

**Tabelle 3: bekannte Kraftfelder und Kraftfeldprogramme**

Kraftfeld	Programm	Hauptanwendungsgebiet
AMBER	AMBER, <sup>[29]</sup> in unterschiedlichen Versionen nach Jahreszahl benannt	Proteine, DNA, Zuckeroligomere, „Biomoleküle“ „sehr große Systeme“
CHARMM	CHARMM, <sup>[30]</sup> in unterschiedlichen Versionen nach Jahreszahl benannt	„Biomoleküle“, „sehr große Systeme“
GROMOS	GROMOS, <sup>[31]</sup> in unterschiedlichen Versionen nach Jahreszahl benannt	„Biomoleküle“, „sehr große Systeme“
MM2	MM2 <sup>[32]</sup>	Organische Moleküle
MM3	MM3 <sup>[33]</sup>	Organische Moleküle, „Biomoleküle“
OPLS	OPLS <sup>[34]</sup>	„Biomoleküle“, Lösemittel, einige organische Moleküle
MMFF94	MMFF94 <sup>[35]</sup>	Parametrisiert an B3LYP/6-31G* Rechnungen, relativ universell
AMBER, MM2, MM3	MACROMODEL <sup>[36]</sup>	Universell, „mittelgroße Systeme“
AMBER, MM+(MM2), MM3, Bio+(CHARMM), OPLS	HYPERCHEM <sup>[37]</sup>	Universell, zusätzlich QM Methoden „kleinere bis mittelgroße Systeme“ da relativ schlecht skalierend
MMX(MM2), MMFF94, AMBER, OPLS	PCMODEL, siehe <a href="http://www.serenasoft.com">www.serenasoft.com</a>	Universell „kleinere bis mittelgroße Systeme“
Tripos <sup>[38]</sup>	SYBYL <sup>[39]</sup>	Organische Moleküle und Proteine, „sehr große Systeme“

Vermehrt werden zudem Moleküldynamik- und Quantenmechanikverfahren auf Teilbereiche des Systems angewandt und gekoppelt. Die im Programm Gaussian03<sup>[2]</sup> implementierte ONIOM Methode ist ein Beispiel.

Die Integration von Gleichung (1.7) erfolgt über kleine Zeitintervalle  $\Delta t$ , typischerweise 1-10 fs. Die Zeitspanne, die simuliert werden kann, liegt wegen des

hohen Rechenaufwands im Nanosekunden- bis Millisekundenbereich, für viele Vorgänge deutlich zu kurz.

Moleküldynamiksimulationen haben einige Vorteile. Sie arbeiten direkt und bei richtiger Wahl der Parameter weitgehend repräsentativ im Phasenraum(1.3). Damit lässt sich nach Equilibrierung des Systems die Moleküldynamiksimulation als Ensemble zur Bestimmung thermodynamischer Größen nach (1.6) direkt heranziehen. Außerdem können dem System ohne Probleme noch weitere Kräfte bzw. Potentiale im Laufe der Simulation zugeführt werden, in (1.9)  $E_{\text{restraint}}$  genannt.

Die meistverbreitete Form für diese zusätzlichen Wirkgrößen ist das harmonische Potential.

$$(1.10) \quad E_{\text{restraint}} = \frac{1}{2} K_{\text{restraint}} (r_{ij} - r_{ij}^0)^2$$

Tabelle 4 zeigt hierfür einige Beispiele:

**Tabelle 4: zusätzliche Kräfte in Kraftfeldern und ihre Verwendung**

zusätzliche Kräfte	Verwendung
Atomfixierung ( <i>coordinate restraints</i> )	Vereinfachung des experimentellen Setups der Rechnung
Abstands-Restraints, Winkel-Restraints, Dieder-Restraints	Beschränkung des Parameterraums, z.B. zur Berücksichtigung von gemessenen NOE-Werten. Nach Kessler <sup>[40]</sup> meist harmonische Potentiale (1.10).
Reibungskräfte und zufällige Kräfte (Kollisionen)	<i>Stochastic Dynamics</i> (SD), verbessertes <i>sampling</i> , Wärmeaustausch mit fixierten Objekten

Der größte Nachteil von MD-Simulationen ist eine recht langsame Abdeckung des verfügbaren Konformationsraums. In Verbindung mit zusätzlichen *restraints* ist dies zugleich ihr größter Vorteil, da die relevanten Teile dementsprechend intensiver untersucht werden. Nach der Röntgenstrukturanalyse sind deshalb MD-Rechnungen in Verbindung mit NOE-Messungen zur zweiten Hauptmethode der Strukturaufklärung in der Biochemie - insbesondere für Proteine und RNA - geworden. Der erste Hauptteil dieser Arbeit verwendet aus diesen Gründen zum großen Teil MD-Rechnungen als *Sample*-Methode der Wahl.

### 1.2.3 Monte-Carlo-Simulationen

Monte-Carlo-Simulationen ähneln den Moleküldynamiksimulationen darin, dass auch sie eine kontinuierliche Trajektorie des Systems im Koordinaten-



raum (1.1) erzeugen. Im Gegensatz zu MD-Rechnungen wird allerdings keine Trajektorie im Phasenraum (1.3) erzeugt. Stattdessen werden in jedem Schritt zufällige Translationen in einem beliebigen Parameterraum durchgeführt. Dies sind meist kartesische Koordinaten, aber auch interne Koordinaten können gewählt werden, um den interessierenden Bereich des Konformationsraums schneller abzuschreiten.

Die aus der vorherigen durch einen Zufallsgenerator geometrisch veränderte Struktur wird bei geringerer Energie immer, bei höherer nur mit einer gewissen Wahrscheinlichkeit akzeptiert. Das meistverwendete Kriterium ist das Metropolis-Kriterium (Gleichung (1.11)).<sup>[41]</sup>

$$(1.11) \quad P_{accept} = \left( \frac{\Delta E_i}{k_B T} \right)$$

Monte-Carlo-Simulationen im kartesischen Koordinatenraum gelten im Vergleich zu MD-Rechnungen als ineffizient.<sup>[42]</sup> Dieses bestätigte sich bei der Implementierung durch das Programm HYPERCHEM in den vorliegenden Versionen,<sup>[37],[43]</sup> weshalb diese Methode hier nicht verwendet wird.

### 1.3 Analyse hochdimensionaler Daten

Wie bereits angesprochen, ergeben sich bei Untersuchungen des Konformationsraums großer Moleküle oft sehr große Datenmengen von hoher Redundanz. Dies ist insbesondere der Fall bei MD- und MC-Rechnungen. Wie aus Gleichung (1.3) unmittelbar hervorgeht, ist der Phasenraum in Systemen mit vielen Freiheitsgraden außerdem von so hoher Dimensionalität, dass eine direkte Darstellung der Ergebnisse zum Scheitern verurteilt ist. Für den menschlichen Betrachter ist beim Betrachten von MD-Trajektorien außer „zappelnden Molekülen“ nicht viel zu sehen.

Zur Analyse hochdimensionaler Daten gibt es in der Literatur verschiedene Ansätze. Einen Überblick über die bekannten Methoden findet sich bei van Gunsteren et al.<sup>[44]</sup> Die einfachste Methode besteht im Ermitteln thermodynamischer Eigenschaften aus den Gesamteigenschaften des Phasenraumes (siehe Gleichung (1.6)) durch numerische Integration. Während dies wichtige und vor allem experimentell verifizierbare Ergebnisse liefert, gehen dabei doch wesentliche strukturelle und dynamische Informationen verloren, so dass diese Methoden in diesem Kapitel nicht bearbeitet werden. Zur Behandlung von strukturellen Daten werden Taxonomien und generell

Ordnungskriterien benötigt, außerdem muss oft in erheblichem Maße Redundanz entfernt werden.

Als Konsequenz wurden in der Literatur Ansätze zur Verwendung klassischer multivarianter Datenanalysemethoden und Dimensionsreduktionstechniken diskutiert. Die wichtigsten sind dabei wohl die Hauptkomponentenanalyse (*principal component analysis (PCA)*),<sup>[45],[46]</sup> sowie die Clusteranalyse (*cluster analysis (CA)*),<sup>[47],[48]</sup> welche bereits in Standardsoftware wie GOPENMOL<sup>[49]</sup> enthalten ist. Im Rahmen der Einleitung werden diese Verfahren kurz vorgestellt, um Vor- und Nachteile gegenüber dem in dieser Arbeit gewählten Verfahren aufzuzeigen.

Die grafische Auftragung interner Koordinaten und direkt erhältlicher Deskriptoren in der *recurrence quantification analysis (RCA)*<sup>[50],[51]</sup> und *multidimensional scaling*<sup>[48]</sup> sind Versuche, MD-Trajektorien zu visualisieren. Neuronale Netze haben bisher nur in Form von ART-2' Netzen Erwähnung in der *Molecular-Modelling-Literatur* gefunden. In den folgenden Unterabschnitten wird ein kurzer Überblick über die einzelnen Verfahren gegeben.

Viele dieser Verfahren verwenden das Konzept der „Ähnlichkeit“. Moleküle gelten dann als ähnlich, wenn ein so genanntes „Ähnlichkeitskriterium“ (*similarity coefficient*) einen hohen Wert annimmt bzw. erfüllt wird, oder ein „Unähnlichkeitskriterium“ (*dissimilarity coefficient*) einen niedrigen Wert annimmt bzw. nicht erfüllt wird. Sind diese Kriterien quantifizierbar, spricht man von „metrischen Verfahren“, welche generell in den Naturwissenschaften vorherrschend sind. Das Prinzip der Ähnlichkeit ist eng verwandt mit dem Begriff der „Nähe“ (*proximity*). Ähnliche Objekte sind sich „nahe“ in Zeit, Raum oder einem anderen Kriterium. Von besonderer Bedeutung ist dabei die „euklidische Distanz“, definiert in Gleichung (1.12)

$$(1.12) \quad d_{euclid}(\mathbf{x}, \mathbf{y}) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

Die genaue Art, in der die Nähe von Objekten beschreibbar ist, wurde von Hartigan<sup>[52]</sup> zur Klassifizierung der anwendbaren Algorithmen herangezogen. Kann die Nähe von Objekten über die euklidische Distanz (1.12) definiert werden, können sehr viele verschiedene statistische Verfahren verwendet werden, unter anderem die in dieser Arbeit verwendeten Kohonen-Netze.

### 1.3.1 Recurrence Analysis (RCA)

Die sogenannte *recurrence analysis* wurde als grafisches Instrument von Eckmann, Kamphorst und Ruelle<sup>[51]</sup> eingeführt, um Muster in Simulationen dynamischer Systeme sichtbar zu machen.

In der einfachsten Form werden die Unterschiede der euklidischen Distanzmatrix  $D_{i,j}$  eines Moleküls aufgetragen. So wird z.B. im Graphen der Abbildung 5 jeder Punkt  $(x, y)$  mit den Funktionswerten aus Gleichung (1.13) eingefärbt.

$$(1.13) \quad f(x, y) = \sum_{i=1}^n \sum_{j=1}^n \sqrt{(D_{i,j}(x) - D_{i,j}(y))^2}$$

Dabei steht  $D_{i,j}(x)$  für den Abstand zwischen Atoms  $i$  und Atom  $j$  zum Zeitpunkt  $x$ , entsprechend für  $D_{i,j}(y)$  zum Zeitpunkt  $y$ .  $x$  und  $y$  sind dabei die Koordinatenachsen, in Abbildung 5  $i$  und  $j$  benannt und in ps angegeben.

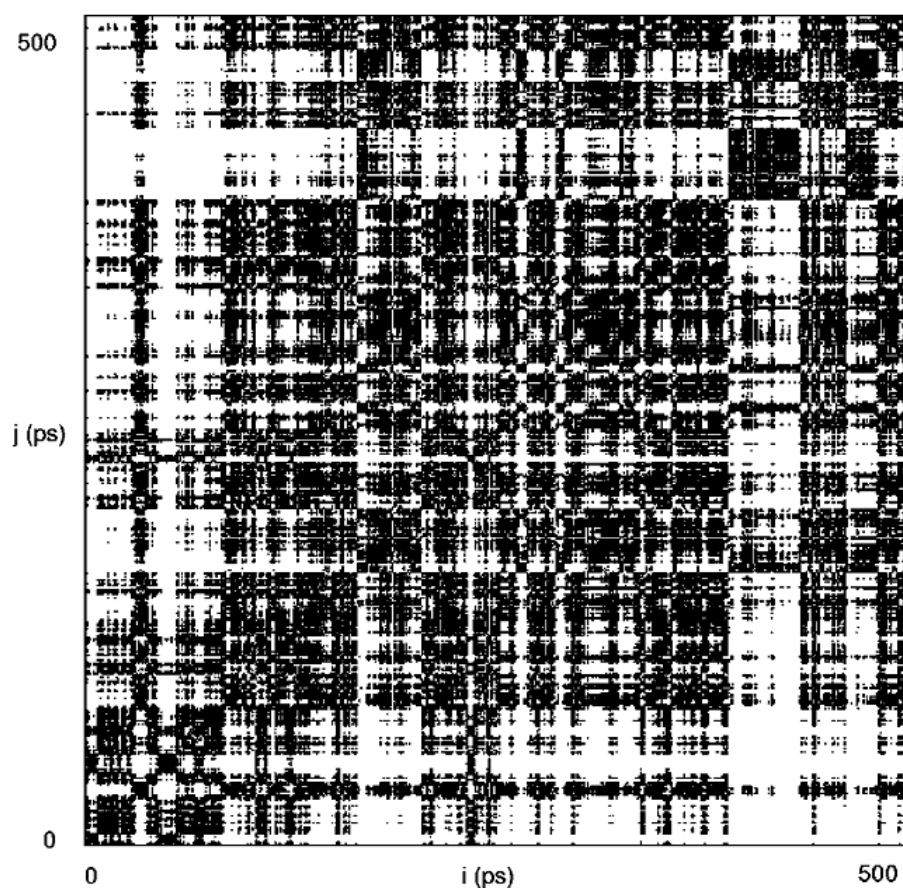


Abbildung 5: **Proteine Recurrence Plot** einer MD-Rechnung eines Proteins, entnommen einer Publikation von Manetti und Ceruso<sup>[50]</sup>

Dunkle Punkte zeigen geringe Unterschiede in der Distanzmatrix und damit hohe Ähnlichkeit an, helle Punkte eine geringe Ähnlichkeit. Das erhaltene Diagramm ist aufgrund der symmetrischen Eigenschaft der Distanzmatrix spiegelsymmetrisch zur Geraden  $x=y$ . Auf diese Art lassen sich Aufenthalte in

einem lokalen Minimum der Hyperfläche, von Manetti und Ceruso „Mikrozustände“ genannt, an dunklen Quadraten um die Gerade  $x=y$  erkennen. „Dynamische Ereignisse“, z.B. Übergang eines Systems aus einem Mikrozustand in den nächsten können an hellen Bereichen um  $x=y$  erkannt werden. Ein Beispiel für die Auswertung einer MD-Rechnung ist in Abbildung 5 aufgeführt.

Die Stärke der *recurrence plots* liegt in der genauen Untersuchung kurzer MD-Trajektorien. Die zeitlichen Informationen bleiben, verknüpft mit konformativen Informationen, erhalten. Bei längeren MD-Simulationen wird der Plot sehr schnell unübersichtlich und verliert an Aussagekraft. Außerdem ist die Einteilung des Graphen in Mikrozustände aufgrund der fraktalen Natur zum Teil recht willkürlich. Eine sehr ähnliche Technik besteht im gezielten Auftragen von Diederwinkeln über die Zeit der Simulation. Die Kombination mehrerer dieser Plots wurde z.B. von Karpen, Tobias und Brooks<sup>[48]</sup> demonstriert und stellt eine reduzierte manuelle Anwendung der RCA-Technik dar, mit ähnlichen Vor- und Nachteilen.

### **1.3.2 Multidimensional Scaling**

Multidimensionale Skalierung dient als Visualisierungsinstrument bei der Darstellung von N-dimensionalen Datensätzen im zweidimensionalen Raum. Eine frühe Anwendung dieser Methode zur Analyse und Visualisierung von MD-Daten stammt von Levitt (1983).<sup>[53]</sup> Nach Karpen, Tobias und Brooks<sup>[48]</sup> sind nur niedrigdimensionale Räume dieser Darstellung zugänglich und beschreiben eine Kombination mit Cluster-Analyse-Techniken. Die einzelnen Konformationen der Simulation werden dabei auf den zweidimensionalen Raum projiziert, so dass ein Gesamtfehler der Projektion minimiert wird. „*Multidimensional Scaling*“ von Cox und Cox<sup>[54]</sup> liefert eine weiter führende Behandlung der verwendeten numerischen Methoden und Anwendungen.

### **1.3.3 Clustering Analysis (CA)**

Die *Clustering*-Verfahren teilen sich auf in zwei große Gruppen: die partitionierenden und die hierarchischen Verfahren.

Partitionierende Verfahren werden auch prototypbasierend genannt. Ziel ist es, im Satz der Daten Clusterzentren zu finden und diesen Nachbarstrukturen zuzuordnen, so dass die Gesamtsumme der euklidischen Distanzen aller zugeordneten Strukturen minimal ist. Es sei  $m_i$  der gemittelte Fehler der euklidischen Distanzen  $x$  über eine Clustermenge (siehe Gleichung (1.14))

$$(1.14) \quad \mathbf{m}_i = \frac{1}{n_i} \sum_{\mathbf{x} \in D_i} \mathbf{x}$$

Die zu optimierende Gesamtsumme der quadratischen Fehler  $J_e$  bei  $c$  Clustern ist dann nach Duda, Hart, Stork<sup>[55]</sup> definiert als Gleichung (1.15).

$$(1.15) \quad J_e = \sum_{i=1}^c \sum_{\mathbf{x} \in D_i} \|\mathbf{x} - \mathbf{m}_i\|^2$$

Die Clusterzentren dienen als Prototypen für die Mitglieder ihres *Clusters*. Damit geht eine massive Redundanzreduktion bei hochredundanten Daten einher, da diese als Statthalter verwendet werden können. Die Zahl der zu suchenden Clusterzentren wird vor der Untersuchung definiert, wodurch es zu falschen Gruppenbildungen bei zu wenigen Clusterzentren kommen kann. Partitionierende Algorithmen haben ihre Stärken in der automatisierten Reduktion von großen Mengen an Datensätzen auf wenige Prototypen. Schwächen sind der Verlust an Ähnlichkeitsbeziehungen zwischen den Clustern und im Falle der Analyse von MD-Rechnungen der Verlust von Informationen über die dynamische Entwicklung des Systems.

Hierarchische Clusterverfahren unterteilen die Menge der zu klassifizierenden Datenpunkte ausgehend von groben Anfangsclustern in immer feinere Untercluster. Es ergeben sich dendritische Datenstrukturen, die sich ausgezeichnet dazu eignen, Beziehungen zwischen einzelnen Konformationen abzubilden. Auch Redundanz stellt kein Problem dar. Während sich hierarchisches *Clustering* zur Bearbeitung von Informationen in automatisierten Verfahren anbietet, neigt es laut Duda, Hart, Stork<sup>[56]</sup> jedoch zu numerischer Instabilität und ist „rauschempfindlich“. Ebenso kommt es zum Verlust von Informationen über die dynamische Entwicklung des Systems.

### **1.3.4 Principal Component Analysis (PCA)**

Im Deutschen auch Hauptkomponentenanalyse genannt, handelt es sich um ein statistisches Verfahren zur Reduktion der Dimensionalität von Eingabedaten. Für  $n$ -dimensionale Daten gibt es grundsätzlich  $n$  Achsen im Koordinatenraum, die orthogonal, d.h. linear unabhängig voneinander sind.

Dies sind bei molekularen Systemen im kartesischen Koordinatenraum in der Regel alle Freiheitsgrade. Nicht alle dieser Achsen tragen jedoch zum Informationsgehalt eines Ensembles an Datenpunkten bei. In der Statistik ist die Varianz das Maß des Informationsgehalts im eindimensionalen Raum. Sie ist definiert in Gleichung (1.16), wobei  $n$  die Gesamtzahl der Datenpunkte ist.

$$(1.16) \quad \text{Var}(X) = s^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{(n-1)}$$

In mehrdimensionalen Räumen gibt die Kovarianz das Maß der linearen Unabhängigkeit zweier Achsen (in statistischer Literatur oft als Variable bezeichnet) an. Sie ist analog zu (1.16) in Gleichung (1.17) definiert.

$$(1.17) \quad \text{Cov}(X, Y) = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{(n-1)}$$

Nur linear unabhängige Achsen tragen zum Gesamtinformationsgehalt bei.

Anstelle der Kovarianz wird für die PCA auch oft die Korrelation zweier Messgrößen verwendet, insbesondere, wenn es sich um Messwerte mit unterschiedlichen Größenbereichen (unskalierte Messwerte) handelt. Diese ist definiert als:

$$(1.18) \quad \text{Kor}(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)} \cdot \sqrt{\text{Var}(Y)}}$$

Der Grundgedanke der PCA besteht darin, eine Hauptachsentransformation so durchzuführen, dass die Korrelation mehrdimensionaler Merkmale für die neue Basis minimiert wird. Anders ausgedrückt, es sollen die Kovarianzen der ersten Achsen möglichst groß und die der späteren so gering wie möglich werden, um Achsen mit kleinen Kovarianzen unter geringstmöglichem Informationsverlust eliminieren zu können. Der komplette Datensatz wird dazu zunächst normiert, so dass  $\bar{\mathbf{x}} = \vec{0}$  gilt. Ist ein Punkt  $\mathbf{x}$  im Koordinatenraum dann durch (1.19) gegeben,

$$(1.19) \quad \mathbf{x} = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}$$

so erhält man die transformierten Datenpunkte  $\mathbf{x}'$  durch die Transformationsgleichung (1.20).

$$(1.20) \quad \mathbf{x}' = \Gamma^T \mathbf{x}$$

$\Gamma$  ist die Matrix der Eigenvektoren  $\gamma_i$  (1.23),  $\Lambda$  die Matrix der Eigenwerte (1.22) in Eigenwertgleichung (1.21)

$$(1.21) \quad \Lambda = \Gamma^T \Sigma \Gamma$$

$$(1.22) \quad \Lambda = \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \lambda_n \end{pmatrix}$$

$$(1.23) \quad \Gamma = \begin{pmatrix} \gamma_{1,1} & \cdots & \gamma_{1,n} \\ \vdots & \ddots & \vdots \\ \gamma_{m,1} & \cdots & \gamma_{m,n} \end{pmatrix}$$

$\Sigma$  ist die symmetrische positiv definite Kovarianzmatrix, definiert in (1.24)

$$(1.24) \quad \Sigma = \begin{pmatrix} \text{Cov}(x_1, x_1) & \cdots & \text{Cov}(x_1, x_n) \\ \vdots & \ddots & \vdots \\ \text{Cov}(x_n, x_1) & \cdots & \text{Cov}(x_n, x_n) \end{pmatrix}$$

Wie bereits beschrieben, kann an dieser Stelle auch die Korrelationsmatrix verwendet werden:

$$(1.25) \quad \Sigma' = \begin{pmatrix} \text{Kor}(x_1, x_1) & \cdots & \text{Kor}(x_1, x_n) \\ \vdots & \ddots & \vdots \\ \text{Kor}(x_n, x_1) & \cdots & \text{Kor}(x_n, x_n) \end{pmatrix}$$

Die PCA ist eine der Analysemethoden, die in Kapitel 3 angewendet werden, um die Hückelaromatizität von Annulenen mit  $4n$   $\pi$ -Elektronen zu belegen.

Prinzipiell bietet sich die PCA dazu an, als Vorstufe der eigentlichen Analyse die Dimensionalität der Eingangsdaten für neuronale Netze und generell *unsupervised-learning*-Methoden zu verringern. Die transformierten Daten können jedoch keinen Eigenschaften des Systems mehr direkt zugeordnet werden. Da im Rahmen dieser Arbeit zum Teil ausführlich von Rückkopplung per Molekülmechanikrechnungen unter Verwendung direkt zuordenbarer Variablen Gebrauch gemacht wird, wird im Rahmen dieser Arbeit auf PCA als Vorstufe der Analyse durch neuronale Netze verzichtet.

### 1.3.5 Neuronale Netze

Wie bereits in Unterkapitel 1.3 angedeutet, gibt es im Bereich *Molecular Modeling* noch nicht viele Anwendungen neuronaler Netze zur Reduktion und Visualisierung der hochkomplexen und redundanten Daten aus Konformationsuntersuchungen und MD-Rechnungen. Bekannt sind hier ART2'-Netzwerke, die z.B. von Brooks *et al.*<sup>[48]</sup> sowie von Grossberg und Carpenter<sup>[57]</sup> beschrieben wurden. Dabei handelt sich um eine clusternde neuronale Netzwerkstruktur, d.h. es handelt sich vielmehr um eine Spezialvariante eines Clusteringverfahrens. Eine weitere Anwendung der Kohonennetze besteht in

der Projektion der dreidimensionalen Struktur eines Moleküls auf die Ebene. Die dreidimensionalen Koordinaten von Atomen des Moleküls dienen dabei als Trainingssatz, propagiert z.B. von Zupan und Gasteiger.<sup>[58]</sup> Die Anwendung mit den dreidimensionalen Positionen der Aminosäuren eines Proteins als Trainingssatz wird von Alper/Bagdassarian berichtet.<sup>[59],[60]</sup> Diese bisherigen Arbeiten haben gemein, dass lediglich statische Moleküle beschrieben werden können. Das heißt, es werden keine Strukturenssembles, sondern nur einzelne Strukturen beschrieben.

Da einer der Hauptpunkte dieser Arbeit die Analyse von Konformerenssembles mittels Kohonennetzen ist, folgt eine etwas ausführlichere Einleitung in das Thema künstliche neuronale Netzwerke in der Informatik. Kohonennetze dienen allgemein zur topologieerhaltenden Projektion aus hohen in niedrigere Koordinatenräume, daher auch der Name *self-organizing feature map*.

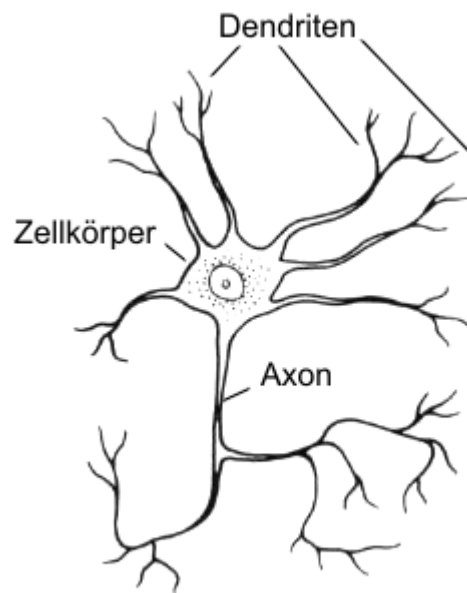
Neuronale Netze bieten die Möglichkeit, Daten zu analysieren und Wissen über ein System aus einem Satz von Daten zu extrahieren. Dies entspricht im Prinzip einer multidimensionalen Regression, für die das theoretische Modell nicht bekannt sein muss und bewirkt eine starke Datenreduktion.

Das Vorbild ist das Gehirn, welches mit seinen Myriaden Nervenzellen in der Lage ist, komplexe Aufgaben zu bewältigen und in einigen Bereichen trotz "langsamster Taktzyklen" jeden Großrechner zu überrunden. Dabei wird kein deterministischer Algorithmus verwendet, sondern die Daten werden von sehr vielen kleinen Datenverarbeitungseinheiten, den Nervenzellen, massiv parallel verarbeitet. Die folgende Nomenklatur wird aus dem Buch „*Neural Network for Chemists*“<sup>[58]</sup> entnommen.

### **1.3.5.1 Neuronale Netze in Biologie und Informatik**

Das menschliche Nervensystem besteht aus ca.  $10^{10}$  Nervenzellen, auch Neuronen genannt. Die mathematische Abstraktion wird im Folgenden anhand von biologischen Modellen erläutert. Da es unterschiedliche Arten von Neuronen gibt, ist in Abbildung 6 der Aufbau eines solchen Neurons stark simplifiziert schematisch dargestellt.

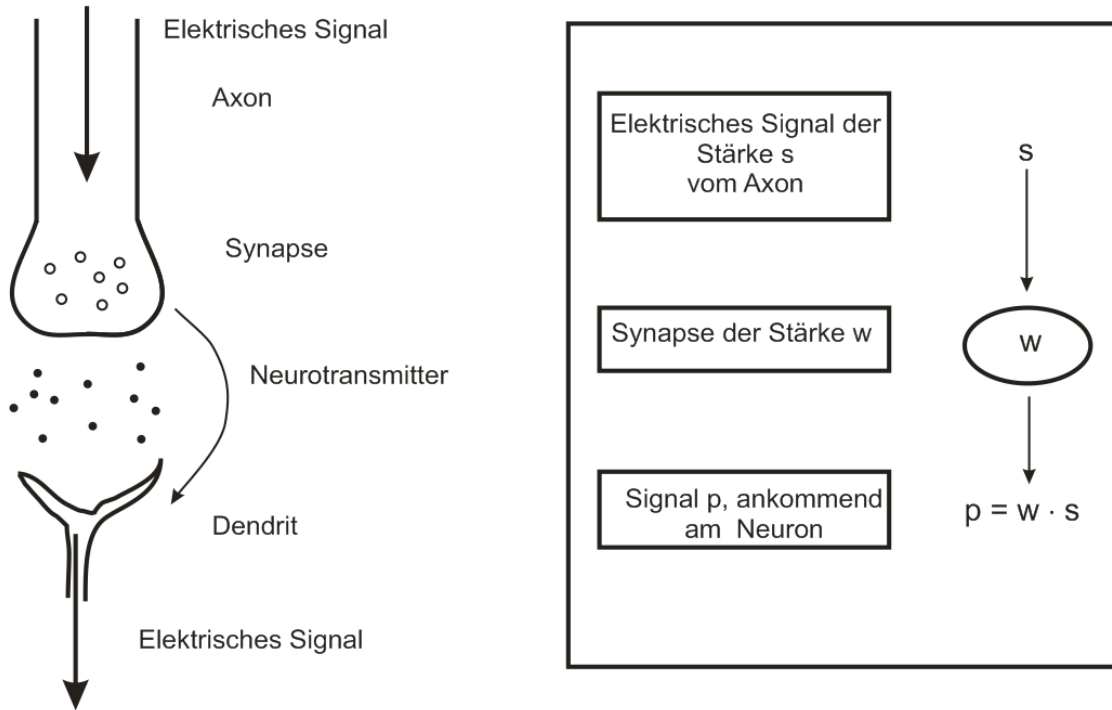




**Abbildung 6:** Schematische Darstellung einer Muskelnervenzelle. Die Verknüpfung für eingehende Signale erfolgt über die Dendriten. Ausgehende Signale werden durch die Axonen übertragen.

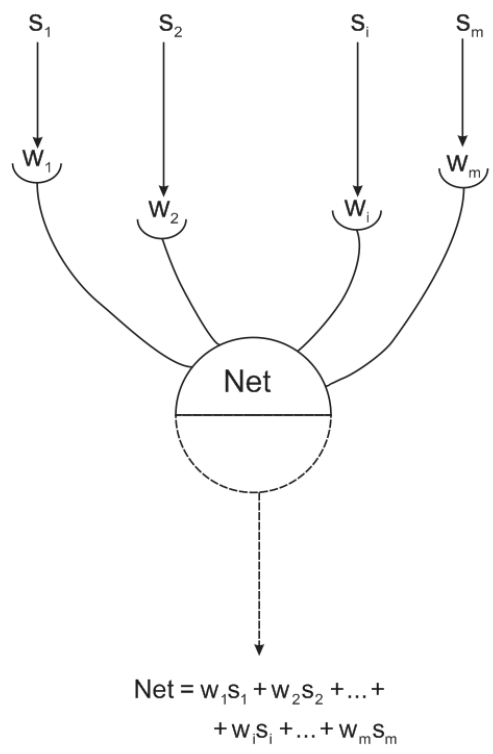
Der Zellkörper hat zwei Arten von Erweiterungen, die Dendriten und das Axon. Die Dendriten "empfangen" Signale von Nachbarneuronen und erzeugen daraus den *Input* für die Neuronen. Das Neuron erzeugt aus dem Input mittels sehr einfacher Funktionen einen *Output*-Wert, der über das Axon, das sich verzweigen kann, via Dendriten an Nachbarneuronen oder Muskelzellen weitergegeben wird.

Die Signalweitergabe vom Axon an den Dendrit wird wie in Abbildung 7 dargestellt mathematisch abstrahiert.



**Abbildung 7:** links die schematisierte Darstellung der Signalübertragung in einer Synapse, rechts die Synapse als *Blackbox Modell*.

Die Verknüpfung aller Eingabewerte in einem einzelnen künstlichen Neuron ist in Abbildung 8 dargestellt.



**Abbildung 8:** Berechnung der Netzeingangsfunktion *Net* an einem künstlichen Neuron

Der Gesamtinput  $Net$  für einen  $m$ -dimensionalen Eingabevektor  $\vec{s}$  an einem Neuron ermittelt sich dann als Vektorprodukt.

$$(1.26) \quad \overrightarrow{Net} = \vec{w} \cdot \vec{s} = w_1 s_1 + w_2 s_2 + \dots + w_i s_i + \dots + w_m s_m$$

Diese einfache Art, den Input für ein Neuron zu berechnen, ist für die meisten Netzwerktypen gültig. Die an das Axon ausgegebene Funktion kann jetzt aus der Netzeingangsfunktion bestimmt werden. Üblich sind einfache Weitergabe der Eingangsfunktion oder vorherige Transformation z.B. mittels einer sigmoidalen Funktion. Das Netzwerk "lernt", indem die Gewichte  $w$  nach bestimmten Regeln an die Eingabedaten angepasst werden.

Der einfachste Fall für ein Netzwerk ist ein einzelnes Neuron, das einen Inputvektor empfängt und dessen Output interpretiert wird.

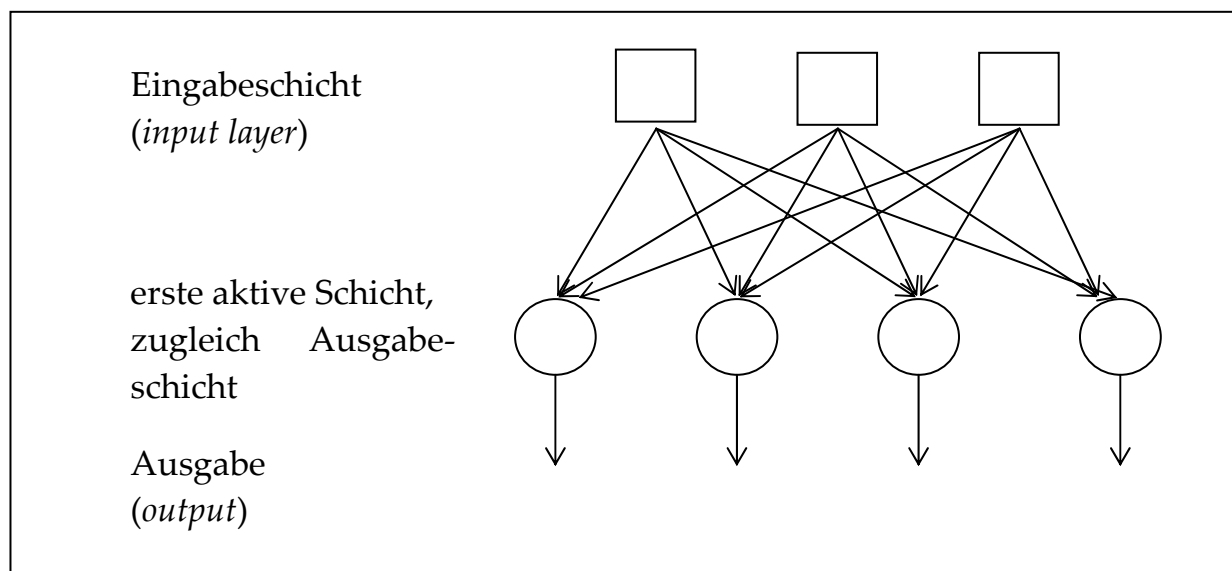
### 1.3.5.2 Netzwerkkonstrukturen

Die Hauptunterschiede zwischen den unterschiedlichen literaturbekannten Netzwerktypen besteht in der Architektur, in der die Neuronen des Netzwerkes miteinander "verdrahtet" sind, sowie in der Art und Weise, in der das Netzwerk lernt, das heißt, die Gewichte angepasst werden.

Üblicherweise werden Neuronen in Schichten (*layers*) angeordnet. Die Eingabedaten ans Netzwerk werden an die erste Schicht gegeben. In der Regel müssen diese Eingabedaten zwischen 0 und 1 normalisiert werden, um sinnvolle Ergebnisse zu erhalten. Der Eingabevektor, das heißt die Schnittstelle zwischen Daten und Netzwerk, entspricht bereits einer ersten Netzwerkschicht, der *input layer*. Dieser Eingabevektor wird gleichzeitig an eine Anzahl Neuronen gegeben, die die erste Schicht oder auch versteckte Schicht (*hidden layer*) bilden. Die Outputwerte all dieser Neuronen können an eine weitere Schicht als Eingabedaten weitergereicht werden. Die letzte Schicht schließlich bildet dann die *output layer*, deren Ausgabe auch die Ausgabe des Netzes genannt wird.

Anhand dieser Charakteristika werden die Netzwerkkonstrukturen einer Nomenklatur unterworfen. In Abbildung 9 handelt es sich danach um ein *one-layer* Design mit einer Eingabeschicht von drei Eingabewerten und einer aktiven Schicht mit vier Neuronen, die zugleich die Ausgabeschicht darstellt.

Zu den bekanntesten Netze gehören z.B. *feedforward networks*, *backpropagation networks*, *Hopfield networks*, *associative neural networks* und als prominentester Vertreter der selbstlernenden Netze die Kohonennetze. Mehr zu letzeterem folgt in Kapitel 2.



**Abbildung 9:** *Feedforward neural network* mit einer Eingabeschicht von drei Elementen, und einer aktiven Schicht von vier Neuronen

## 1.4 Aromatizität

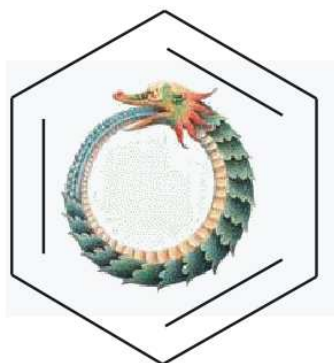
Die Untersuchung über die Aromatizität von Annulenen in Möbius- und Hückelkonfiguration ist Thema des zweiten Hauptabschnitts dieser Arbeit. Dabei kommt es zu zwei Komplikationen, die eine Anwendung der klassischen Methoden des *Molecular Modellings* schwierig machen. Zum einen zeigen die untersuchten Moleküle eine komplexe Hyperfläche mit vielen Konfigurations- und Konformationsminima, die es notwendig machen, einen größeren Satz Strukturen zu berücksichtigen. Dies erfordert die Durchführung einer Konformationssuche als ersten Schritt. Die vorgestellte Zielsetzung beinhaltet dabei ausdrücklich die Untersuchung elektronischer Effekte. Damit kann nur sehr bedingt auf molekülmechanische Methoden zurückgegriffen werden. Eine effiziente Untersuchung des Konformationsraums und eine Voruntersuchung mittels der im Rahmen dieser Arbeit entwickelten CANE - Deskriptoren (CANE: *canonical annulene enumeration*) dienen dazu, die Redundanz und damit die extreme Rechenlast der dichtefunktionalen Methoden zu reduzieren.

Zum zweiten ist der Begriff der Aromatizität nicht präzise definiert. Im Sinne der Quantenmechanik handelt es sich nicht um eine Observable. Das heißt, es handelt sich nicht um eine direkt messbare Eigenschaft, sondern um ein Konzept, welches durch empirische Interpretation von Observablen (Geometrie, Energie, magnetische Eigenschaften) abgeleitet wird. Nichtsdestoweniger handelt es sich um ein Konzept, das ähnlich den Begriffen „weich“

und „hart“ in der chemischen Literatur anerkannt und von praktischem Nutzen ist.

### 1.4.1 Aromatizität als historischer Begriff

Als aromatisch bezeichnete Moleküle wurden in der Chemie relativ schnell als eigenständige „Klasse“ mit gemeinsamen Eigenschaften und Reaktivitäten erkannt. Die erste, von M. Faraday 1825 isolierte<sup>[61]</sup> aromatische Verbindung und zugleich sicherlich die bekannteste, ist das Benzol. Die Struktur dieses Moleküls war den Chemikern zur damaligen Zeit ein Rätsel. So war zwar bekannt, dass das Molekül aufgrund der elementaren Zusammensetzung ungesättigt sein musste, aber nicht die Reaktionen einging, die diese Verbindungsklasse normalerweise eingeht. Außerdem wurden weniger Isomere beobachtet, als bei Annahme von lokalisierten Doppelbindungen eigentlich erwartet wurden. Angeblich in einem Traum über eine Schlange, die sich selbst in den Schwanz beißt, kam August Friedrich Kekulé<sup>[62]</sup> auf die nach ihm benannte ringförmige Struktur.



**Abbildung 10: Kekulé's Traumstruktur**

Hauptmerkmal der Aromatenchemie ist die unerwartet hohe Stabilität der ungesättigten Kohlenstoff-Kohlenstoff-Doppelbindungen. Additionsreaktionen am Aromatenkern gelingen normalerweise nicht, überwiegend herrschen Substitutionsreaktionen vor. Nachdem 1926 Schrödinger die nach ihm benannte Gleichung eingeführt hatte, hielt die Quantenmechanik auch Einzug in die Chemie. Hückel wandte die Theorie als erster konsequent unter starken Vereinfachungen auf konjugierte  $\pi$ -Systeme an und leitete daraus die  $4n+2$  Regel für Aromaten ab.<sup>[63]</sup> Entsprechend gelten planare  $[4n]$ Annulene als antiaromatisch. Antiaromatische Substanzen sind in der Regel schwer zu synthetisieren und weichen der Hückelantiaromatizität durch Einnehmen einer nicht planaren oder stark lokalisierten Struktur aus. Heilbronner postulierte analog dazu  $[4n]$ Annulene in Möbius-Konfiguration als aromatisch und ihre  $[4n+2]$  Gegenstücke als antiaromatisch.<sup>[64]</sup>

Die außergewöhnliche Stabilität der Aromaten wurde bald in „energetischen Kriterien“ quantifiziert. Der Vergleich von Verbindungen mit normalen konjugierten Doppelbindungen und ihren aromatischen Gegenstücken liefert dabei eine aromatische Stabilisierungsenergie (*aromatic stabilization energy* - ASE), auch Resonanzenergie genannt. Z.B. beim Benzol im Vergleich zum hypothetischen Cyclohexatrien (extrapoliert aus den experimentellen Hydrierwärmern von Benzol und Cyclohexen) nimmt diese den beträchtlichen Wert von 36 kcal/mol an.<sup>[65]</sup> Es gibt verschiedene konkurrierende Verfahren, thermodynamische Kreisprozesse aufzustellen. Im Rahmen dieser Arbeit widmen wir uns dabei besonders dem ISE-Index<sup>[66]</sup> (*isomerization stabilization energy*), bzw. einer Abart, dem ISE<sub>π</sub>-Index (*indene isoindene ISE*) und einer selbstentwickelten Modifikation, dem FISE-Index (*fixed indene isoindene stabilization energy*).

In der Struktur einer Verbindung spiegelt sich Aromatizität durch einen Bindungslängenausgleich zwischen Einfach- und Doppelbindungen der Resonanzstrukturen wieder. So gibt es z.B. in Benzol nur eine einzige C-C Bindungslänge. Aromatizitätskriterien, die von der Struktur eines Moleküls abgeleitet werden, sind z.B. HOMA<sup>[67]</sup> und Julg-Index.<sup>[68]</sup>

Eine weitere Klasse an Kriterien ergibt sich aus dem spezifischen Verhalten von aromatischen Verbindungen im Magnetfeld. Sowohl die eigentümlichen Werte der magnetischen Suszeptibilität<sup>[69],[70],[71]</sup> als auch deutliche Verschiebungen in NMR-Spektren (*nuclear magnetic resonance*) werden durch Paulings Ringstromtheorie erklärt.<sup>[72]</sup> Auf quantenmechanischen Berechnungen von NMR-Verschiebungen baut Schleyers NICS (*nucleus independent chemical shift*) auf,<sup>[73],[74]</sup> mehr dazu im Unterkapitel 1.4.4.

Des Weiteren sind Verallgemeinerungen beschrieben worden:

- das Aromatizitätsprinzip in Übergangszuständen<sup>[75],[76],[77]</sup> von Zimmerman,
- Aromatizität in nicht ringförmigen, konjugierten planaren Systemen (Clarsche Regel für aromatische Sextetts<sup>[78]</sup>),
- dreidimensionale aromatische Strukturen, z.B. nichtklassische Kationen.

Während all diese Kriterien dazu verwendet werden können, um Aromatizität numerisch zu definieren, so fehlte doch lange Zeit eine allgemeingültige Methode zur graphischen Repräsentation von Aromatizität. Eine ebensolche wurde von Herges und Geuenich in Form der ACID-Methode (*Anisotropy of the Induced Current Density*)<sup>[79],[80]</sup> vorgestellt und soll zur Visualisierung und Überprüfung der Ergebnisse in Kapitel 3 dienen.

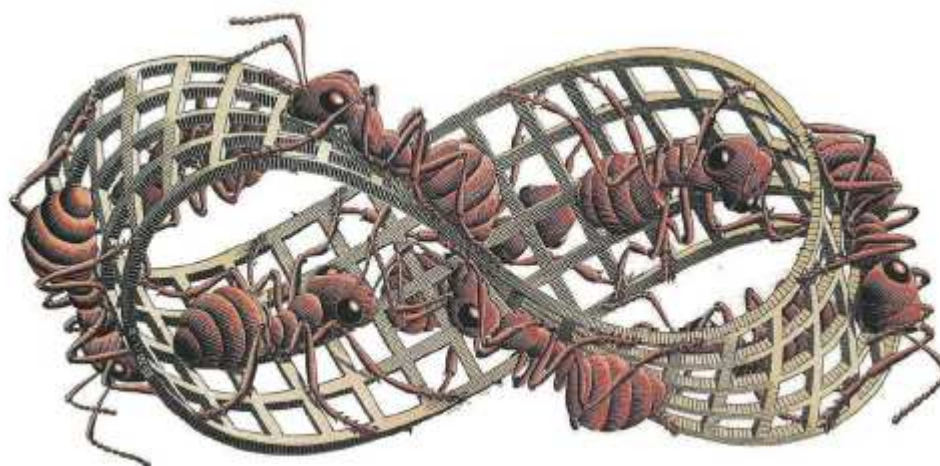
Einen Überblick über einige wichtige Meilensteine in der Entwicklung des Prinzips der Aromatizität vermittelt Tabelle 5.

**Tabelle 5: Aromatizität - historischer Abriss**

1825	Isolierung des Benzols durch Faraday <sup>[61]</sup>
1865	Benzolstruktur, Kekulé <sup>[62]</sup>
1866	Aromatisches Substitutionsprinzip, Erlenmeyer <sup>[81]</sup>
1910	Diamagnetische Suszeptibilität in Aromaten, Pascal <sup>[82]</sup>
1931	Hückeltheorie ( $4n+2$ ) <sup>[63]</sup>
1936	Ringstromtheorie, Pauling <sup>[72]</sup>
1964	Möbius-Aromaten, Heilbronner <sup>[64]</sup>
1968	Diamagnetismus in Aromaten, Dauben <sup>[69]</sup>
1980	Erste quantenchemische Rechnungen über magnetische Eigenschaften von Molekülen mit dem Programm IGLO, Kutzelnigg <sup>[83]</sup>
1996	NICS Nucleus independent chemical shifts, Schleyer <sup>[74]</sup>
2001	ACID Anisotropy of the induced current density, Herges <sup>[79],[80]</sup>
2003	Synthese der ersten durch Röntgenstrukturanalyse bestätigten möbiusaromatischen Verbindung, Herges/Ajami <sup>[84]</sup>

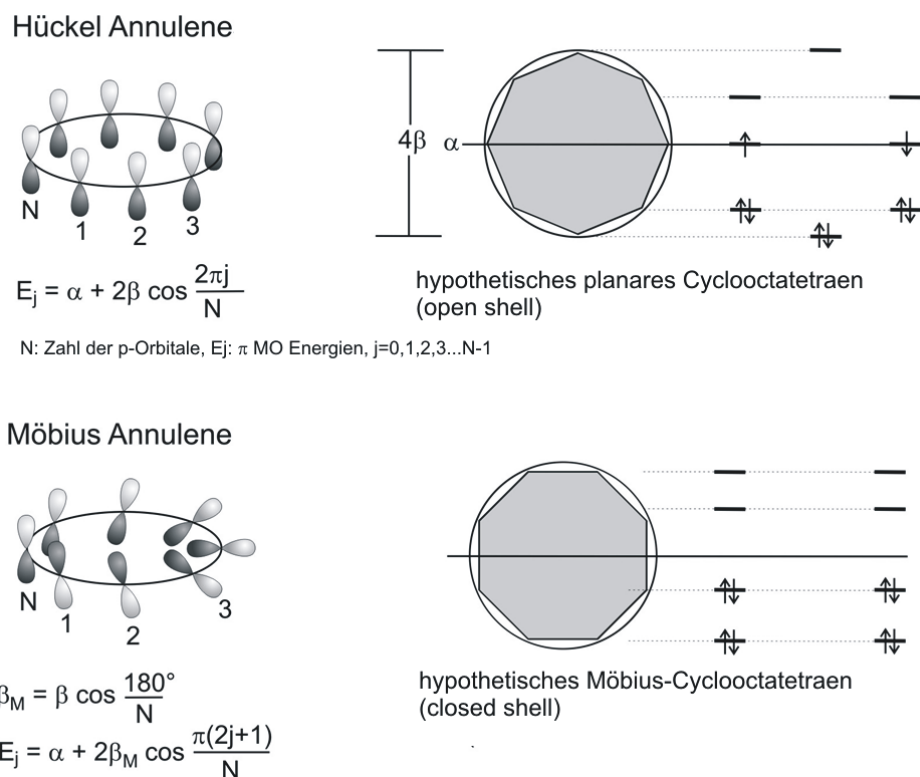
#### 1.4.2 Hückel- und Möbiusaromatizität in Annulenen

Das seltsame topologische Konstrukt einer einseitigen Oberfläche, das heute unter dem Namen Möbius-Band bekannt ist, wurde 1858 unabhängig voneinander von Listing und Möbius<sup>[85]</sup> zum ersten Mal mathematisch beschrieben. Mit Papier und Schere erhält man ein Möbius-Band, indem man einen rechteckigen Streifen Papier einmal um  $180^\circ$  verdrillt und ihn dann an den Enden zum Ring verklebt. Eine besonders ästhetische Darstellung findet sich in den Werken von M.C. Escher (siehe Abbildung 11).



**Abbildung 11: M.C. Eschers Holzschnitt eines Möbiusbandes, auf dem eine Ameisenstraße ihre unendliche Reise antritt**

Heilbronner erkannte 1964,<sup>[64]</sup> dass  $[4n]$ Annulene, die normalerweise nach der Hückel-Regel antiaromatisch und instabil wären, in verdrehter Möbius-Topologie potentiell aromatisch sind. Anschaulich dargestellt wird das Energieschema der  $\pi$ -Orbitale in Abbildung 12.



**Abbildung 12: Frost-Musulin Mnemonik zur Bestimmung der MO-Energien (Molekülorbital-Energien) in Hückel- und Möbiusannulenen**

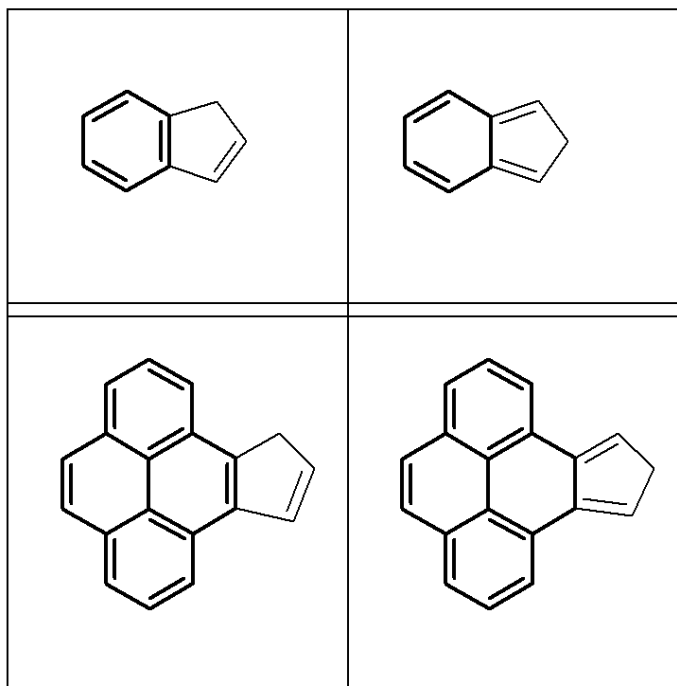
Da einige Bedingungen der Hückelnäherung (Separation der  $\sigma$ - und  $\pi$ -Orbitale) in den notwendigerweise nichtplanaren Möbiussystemen nicht mehr gelten, ist jedoch prinzipiell fraglich, ob jenseits der einfachen Hückelnäherung makroskopisch Eigenschaften beobachtet werden können, die auf eine Möbiusaromatizität hinweisen. Eine Untersuchung des Isomerenensembles einiger  $[4n]$ -Annulene soll im Rahmen dieser Arbeit Aufschluss darüber geben, ob Möbius-Annulene zumindest rechnerisch typisch aromatische Eigenschaften zeigen.

### 1.4.3 Energiekriterien: $ISE_{\pi}$ , $FISE$

Erstmals von Schleyer 2003 vorgeschlagen,<sup>[66]</sup> handelt es sich beim  $ISE_{\pi}$ -Wert um eine Methode, die aromatische Stabilisierungsenergie (ASE) eines konjugierten Systems auf systematische Weise zu bestimmen.

Der „Trick“ liegt dabei darin, nicht das Molekül selber, sondern die cyclopentadien-annelierten Ringsysteme als inden- und isoinden-analoge Isomere zu betrachten, siehe Abbildung 13.





**Abbildung 13:** Inden- und Isoindenderivate des Benzols (oben) sowie des Pyrens (unten). Die tatsächlichen Molekülgerüste vor der Derivatisierung sind fett dargestellt.

Beide Moleküle haben die gleiche Anzahl der Atome, Atomtypen und Bindungen. Beim Isoindenderivat kommt es jedoch zu einer Unterbrechung der Konjugation des  $\pi$ -Systems, während diese beim Indenderivat erhalten bleibt.

Die Stabilisierungsenergie ISE berechnet sich damit zu

$$(1.27) \quad ISE_{II} = \Delta H(\text{Isoinden}) - \Delta H(\text{Inden})$$

Schleyer schlägt außerdem die Verwendung von *syn-anti*-Korrekturen vor und erhält damit für Benzol folgende Stabilisierungsenergien. (Tabelle 6)

**Tabelle 6:** aromatische Stabilisierungsenergien des Benzols aus experimentellen Daten nach Schleyer<sup>[66]</sup>

ISE <sub>II</sub>	21.8 kcal/mol
ISE <sub>IIcorr</sub>	29.0 kcal/mol

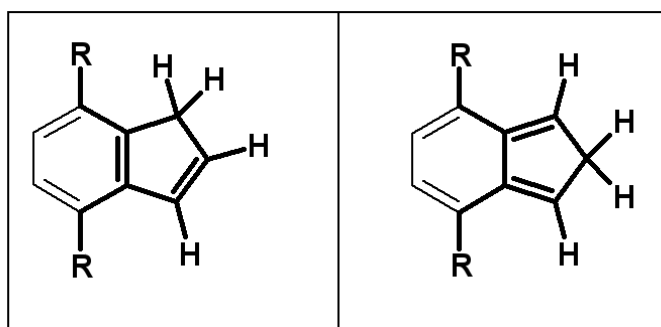
Im Rahmen dieser Arbeit wird der ISE<sub>II</sub> Wert unkorrigiert verwendet, da grundsätzlich keine absoluten, sondern relative Werte behandelt werden. In diesem Fall egalieren sich die Korrekturen und man erhält das gleiche Ergebnis wie beim Verwenden der korrigierten Version.

Prinzipiell lässt sich der ISE<sub>II</sub>-Wert für alle cisoiden, sterisch nicht zu überfrachteten Doppelbindungen bestimmen. Interessanterweise handelt es sich beim ISE-Verfahren eigentlich um einen Wert, der allgemein als Konjugations-Stabilisierungsenergie (CSE) beschrieben werden kann. Er beschreibt die

zusätzliche energetische Stabilisierung, die eine Doppelbindung als Konjugationsbrücke des restlichen  $\pi$ -Systems erfährt. Auch wenn das ISE-Verfahren bisher fast ausschließlich Anwendung auf aromatische und antiaromatische Ringsysteme gefunden hat,<sup>[66]</sup> sollte nicht vergessen werden, dass es sich bei der ASE nur um einen Spezialfall der CSE handelt. So sind bei den folgenden Beispielen, siehe Abbildung 15, sowohl Aromaten, Antiaromaten und konjugierte Polyene aufgeführt.

Am Stammsystem wird die in Abbildung 13 beschriebene Derivatisierung durchgeführt und nach quantenmechanischer Optimierung der  $ISE_{II}$  Wert berechnet. Allerdings wurde im Rahmen der Untersuchungen die Erfahrung gemacht, dass aromatische Möbiusisomere der  $[n]$ Annulene diese Prozedur nur schwer tolerieren. Es kommt im Verlauf der Optimierung mit großer Wahrscheinlichkeit zu drastischen Strukturveränderungen im Molekül, wobei sehr oft auch die Topologie von möbiusartiger zu hückelartiger Struktur wechselt. Damit ist die Methode in der Form für die vorliegende Aufgabe unbrauchbar. Eine quantenmechanisch nicht ganz konsistente Modifikation des Verfahrens, FISE (*fixed ISE*), erweist sich jedoch als deutlich robuster und durchaus aussagekräftig.

Dabei wird nach der Derivatisierung ein großer Teil des Moleküls im Verlauf der Optimierung festgehalten. Dies ist grafisch anhand von Abbildung 14 dargestellt.



**Abbildung 14:** FISE Methode: Die zu optimierenden Atome sind fett dargestellt. Alle übrigen Atome werden im Verlauf der Optimierung festgehalten (*fixed*). Der Rest R wird nur optimiert, wenn es sich um ein Wasserstoffatom handelt, also  $R=H$ .

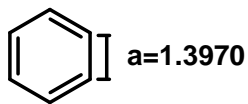
Inwieweit eine solche Vereinfachung wirklich zulässig ist, soll ein Vergleich mit Schleyers  $ISE_{II}$ -Verfahren zeigen. Die Strukturen in Abbildung 15 werden auf B3LYP/6-31G\* Niveau<sup>[86],[87]</sup> optimiert, substituierbare Bindungen sind bezeichnet und mit Längenangaben versehen. Als substituierbar werden in diesem Fall alle cisoiden Bindungen zwischen zwei  $sp^2$ -Kohlenstoffatomen

angesehen, welche zwei geminale Wasserstoffatome besitzen, auch wenn das Verfahren eigentlich nur für Doppelbindungen und nicht für Einfachbindungen chemisch sinnvolle Strukturen ergeben sollte. Eine eindeutige Klassifikation als Doppel- und Einfachbindung ist allerdings aufgrund des Bindungslängenausgleichs und der mesomeren Grenzformen bei den angegebenen Beispielen in Abbildung 15 nicht möglich, sämtliche der Bindungen können auch als Doppelbindungen gezeichnet werden.

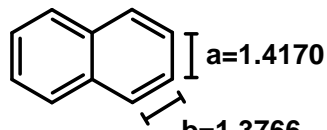
Die ebenfalls auf B3LYP/6-31G\*-Niveau erhaltenen Energien der optimierten Inden/Isoinden-Derivate sowie die ISE<sub>II</sub>- und FISE-Werte sind in Tabelle 7 wiedergegeben.

polyzyklische Aromaten:

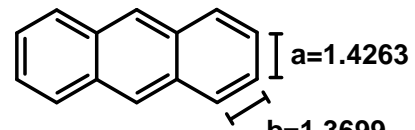
Benzol

 $D_{6h}$ 

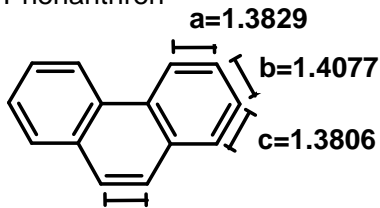
Naphthalen

 $D_{2h}$ 

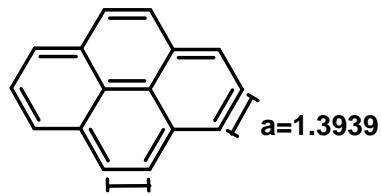
Anthracen

 $D_{2h}$ 

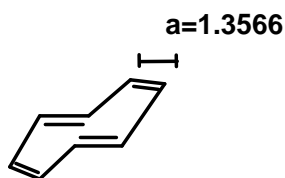
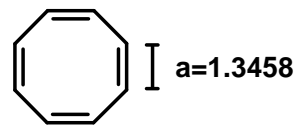
Phenanthren

 $C_{2v}$ 

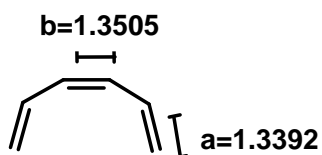
Pyren

 $D_{2h}$ Antiaromaten:

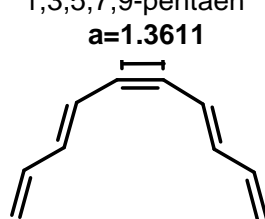
(1E,3Z,5E,7Z)-Cyclooctatetraen

 $C_{2h}$ hypothetisches planares  
(1Z,2Z,4Z,6Z)-Cyclooctatetraen $D_{4h}$ konjugierte Polyene:

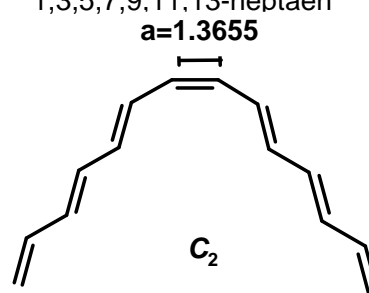
(Z)-Hexa-1,3,5-trien

 $C_2$ 

(3E,5Z,7E)-Deca-1,3,5,7,9-pentaen

 $C_2$ 

(3E,5E,7Z,9E,11E)-Tetradeca-1,3,5,7,9,11,13-heptaen

 $C_2$ 

(3E,5E,7E,9Z,11E,13E,15E)-Octadeca-1,3,5,7,9,11,13,15,17-nonaen

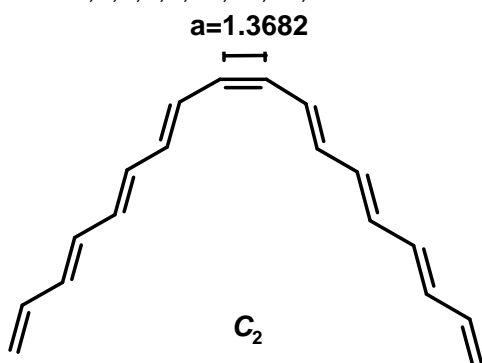
 $C_2$ Abbildung 15: Beispielmoleküle für die ISE<sub>II</sub>- und FISE-Rechnungen in Tabelle 7, Bindungslängen in Å

Tabelle 7: ISE<sub>II</sub>- und FISE-Energien der Beispiele aus Abbildung 15, Nullpunktsenergien in Hartree, (F)ISE-Energien in kcal/mol

Struktur / Derivatisierung	<i>nonfixed derivatives</i> (ISE)		<i>fixed derivatives</i> (FISE)		ISE <sub>II</sub>	FISE
	E <sub>0</sub> (Ind)	E <sub>0</sub> (Isoind)	E <sub>0</sub> (Ind)	E <sub>0</sub> (Isoind)		
<b>polyzyklische Aromaten:</b>						
Benzol, <i>D</i> <sub>6h</sub>						
a	-347.62192	-347.58724	-347.62190	-347.58486	21.76	23.25
Naphthalen, <i>D</i> <sub>2h</sub>						
a	-501.21968	-501.17027	-501.21964	-501.16732	31.01	32.83
b	-501.21817	-501.19662	-501.21799	-501.19457	13.53	14.70
Anthracen, <i>D</i> <sub>2h</sub>						
a	-654.81133	-654.75494	-654.81129	-654.75184	35.39	37.30
b	-654.80952	-654.79370	-654.80933	-654.79215	9.93	10.78
Phenanthren, <i>C</i> <sub>2v</sub>						
a	-654.81349	-654.78584	-654.81232	-654.78126	17.35	19.49
b	-654.81823	-654.77523	-654.81818	-654.77242	26.98	28.71
c	-654.81677	-654.79079	-654.81663	-654.78858	16.30	17.61
d	-654.81563	-654.79967	-654.81507	-654.79690	10.02	11.40
Pyren, <i>D</i> <sub>2h</sub>						
a	-731.03958	-731.00622	-731.03945	-731.00391	20.93	22.30
b	-731.03793	-731.02265	-731.03746	-731.02042	9.59	10.69
<b>Antiaromaten:</b>						
(1E,3Z,5E,7Z)-Cyclooctatetraen, <i>C</i> <sub>2h</sub>						
a	-424.83864	-424.84503	-424.83817	-424.84492	-4.00	-4.24
hypothetisches planares (1Z,2Z,4Z,6Z)-Cyclooctatetraen, <i>D</i> <sub>4h</sub>						
a	-424.91510	-424.93125	-424.91510	-424.93006	-10.13	-9.39
<b>konjugierte Polyene:</b>						
(Z)-Hexa-1,3,5-trien, <i>C</i> <sub>2</sub>						
a	-348.73528	-348.73836	-348.73528	-348.73816	-1.94	-1.80
b	-348.73923	-348.73932	-348.73893	-348.73862	-0.06	0.19
(3E,5Z,7E)-Deca- 1,3,5,7,9-pentaen, <i>C</i> <sub>2</sub>						
a	-503.48647	-503.48459	-503.48644	-503.48410	1.18	1.47
(3E,5E,7Z,9E,11E)-Tetradeca- 1,3,5,7,9,11,13-heptaen, <i>C</i> <sub>2</sub>						
a	-658.23436	-658.23156	-658.23430	-658.23087	1.76	2.15
(3E,5E,7E,9Z,11E,13E,15E)-Octadeca-1,3,5,7,9,11,13,15,17-nonaen, <i>C</i> <sub>2</sub>						
a	-812.98262	-812.97918	-812.98257	-812.97842	2.16	2.61

Die absoluten Zahlen der FISE-Energien liegen systematisch um ca. 1 kcal/mol höher als ihre ISE<sub>II</sub>-Gegenstücke und zwar unabhängig von der Substitutionsposition und Art des Aromaten (aromatisch, nicht aromatisch, antiaromatisch,

hoher Doppel- oder Einfachbindungscharakter). Damit befinden sich die Werte im Rahmen der Genauigkeit der DFT-Rechnungen in exzellenter Übereinstimmung und die FISE-Methode kann qualitativ als äquivalent zur ISE<sub>II</sub>-Methode angesehen werden, obwohl ein Großteil der Atome nicht mitoptimiert wird.

Im Falle der aromatischen Ringsysteme kann der ISE bzw. FISE-Wert als aromatische Stabilisierungsenergie betrachtet werden. Dabei gibt es im gleichen Molekül, ja sogar im gleichen Ringsystem stark unterschiedliche Werte, wie zum Beispiel beim Naphthalin an den a- und b-Positionen, siehe Abbildung 15 und Tabelle 7. Offensichtlich spielt hier das erhaltene elektronische System bei der Isoinden-Derivatisierung eine entscheidende Rolle. Im Falle von Doppelbindung a verbleibt ein radialenartiges Molekül, an Position b ein intakter Benzolring. Auf die Frage: „Wie aromatisch ist ein Ring?“ kann die Konjugations-Stabilisierungsenergie also ganz offensichtlich nicht immer einen eindeutigen Wert liefern. Diese Problematik wird im übrigen von den Autoren der ISE-Methode<sup>[66]</sup> elegant ausgeklammert, indem ausschließlich hochsymmetrische planare Annulene mit einer einzigen substituierbaren Doppelbindung von ihnen betrachtet werden. Diese potentielle Mehrdeutigkeit der erhaltenen Ergebnisse stellt sicher einen berechtigten Kritikpunkt dar, der allerdings von den übrigen thermodynamische Kreisprozessen zum Erhalt einer ASE geteilt wird. Trotzdem liefern Energiekriterien vom Trend her richtige und gut interpretierbare Daten.

Aromaten zeigen eine Stabilisierungsenergie von 10-40 kcal/mol, Antiaromaten eine hypothetische Destabilisierung von -10 kcal/mol ( $D_{4h}$ -Cyclooctatetraen). De facto weicht die Struktur der antiaromatischen Destabilisierung jedoch weitgehend durch Einnahme einer nichtplanaren Struktur aus ( $C_{2h}$ -Cyclooctatetraen). Dabei ist die Überlappung des  $\pi$ -Systems vermindert und man erhält einen immer noch negativen Wert von -4 kcal/mol für das  $C_{2h}$ -Cyclooctatetraen. Für lineare konjugierte Polyene erhält man erwartungsgemäß ISE<sub>II</sub>- und FISE-Werte von um die 0 kcal/mol. Die CSE steigt bei den linearen Polyenen asymptotisch mit der Gesamtanzahl an konjugierten Doppelbindungen auf Werte  $\geq 2$  kcal/mol leicht an. Auch nichtaromatische konjugierte Systeme können eine positive CSE (sprich ISE- und FISE-Werte) haben.

#### **1.4.4 Magnetische Kriterien: NICS, MAGS, FISENICS, FISEMAGS**

Magnetische Eigenschaften haben sich als Aromatizitätskriterien vor allen Dingen aus zwei Gründen als nützlich erwiesen. Zum einen sind experimen-

telle Daten gut erhältlich. Die diamagnetische Suszeptibilität  $\chi$  wurde, obwohl deutlich schwächer als paramagnetische Effekte, schon recht früh beobachtet. Ebenso ist die aromatische Verschiebung in NMR-Spektren ein markanter Effekt und leicht zu messen.

Zum anderen sind magnetische Eigenschaften wie NMR-shift und magnetische Suszeptibilität (MAGS) rechnerisch ausgesprochen gut zugänglich. Die magnetische Suszeptibilität (MAGS) kann beim Vergleich zweier identischer Moleküle in unterschiedlichen Konformationen direkt als Aromatizitätskriterium verwendet werden. Ansonsten wird oft die Differenz  $\Lambda$  aus dem Messwert  $\chi_m$  (auch quantenmechanisch berechnete Werte gelten in diesem Sinne als Messwerte) und dem aus einem Inkrementsystem berechneten hypothetischen Wert  $\chi_a$  herangezogen (*magnetic susceptibility exaltation*).<sup>[88]</sup>

$$(1.28) \quad \Lambda = \chi_m - \chi_a$$

Diese Arbeit verwendet ausschließlich das erste Verfahren.

Der NICS(0) Wert ist definiert<sup>[73],[66]</sup> als das Negative der magnetischen Abschirmung im Zentrum eines zyklisch konjugierten Systems. Praktisch wird dabei eine NMR-GIAO Analyse durchgeführt, wobei sich ein *ghost-atom* an der NICS-Position befindet. Eine angebliche Verbesserung des Verfahrens nach Schleyer, der NICS(1)-Wert, geht vom Wert ein Ångström über der Position des NICS(0)-Wertes aus, wird aber in der Literatur praktisch nicht verwendet, so auch nicht in dieser Arbeit.

Es erweist sich in dieser Arbeit, dass NICS- und MAGS-Werte in ihrer absoluten Form nur bedingt geeignet sind, Strukturen als aromatisch oder nichtaromatisch zu klassifizieren. Es wird an dieser Stelle ein Ansatz vorgeschlagen und untersucht, der anhand des Vorzeichens direkt die Aromatizität oder Antiaromatizität eines Annulens belegen soll. Dazu werden die NICS und MAGS-Werte der im ISE-, bzw. FISE-Verfahren erhaltenen Strukturen herangezogen, um aus ihnen die Differenz und damit ISENICS-, ISEMAGS, bzw. die für automatisierte Verarbeitung besser geeigneten FISENICS- und FISEMAGS-Deskriptoren zu ermitteln.

$$(1.29) \quad ISENICS = NICS(\text{Indenderivat}) - NICS(\text{Isoindenderivat})$$

$$(1.30) \quad ISEMAGS = MAGS(\text{Indenderivat}) - MAGS(\text{Isoindenderivat})$$

$$(1.31) \quad FISENICS = NICS(\text{fixed Indenderivat}) - NICS(\text{fixed Isoindenderivat})$$

$$(1.32) \quad FISEMAGS = MAGS(\text{fixed Indenderivat}) - MAGS(\text{fixed Isoindenderivat})$$

### 1.4.5 Strukturelle Kriterien: HOMA, Julg-Index, GEO, EN

Strukturelle Aromatizitätskriterien haben sich vor allem in statistischen Analysen als robuste Indikatoren der Aromatizität erwiesen (Katritzky 1989).<sup>[89]</sup> Die folgenden, im Rahmen dieser Arbeit im Programm XYZ2CANE verwendeten Definitionen und Gleichungen, halten sich an das Schema von Krygowski und Cyranski.<sup>[90]</sup>

Generell sprechen aus struktureller Sicht Bindungslängenausgleich und Bindungslängenvergrößerung für Aromatizität. Ein ausschließlich auf Bindungslängenausgleich basierender Index ist der Julg-Index<sup>[68]</sup> (siehe (1.33)),

$$(1.33) \quad A_J = 1 - \frac{225}{n} \sum_{r=1}^n \left(1 - \frac{R_r}{R}\right)^2 : \text{Julg-Index}$$

wobei  $n$  die Anzahl der konjugierten Bindungen,  $R_r$  den individuellen Bindungslängenabstand und  $R$  den mittleren Bindungslängenabstand bezeichnet.

Ein weiterer gebräuchlicher struktureller Index ist der HOMA-Index (*Harmonic Oscillator Model of Aromaticity*).<sup>[91],[92]</sup>

$$(1.34) \quad HOMA = 1 - \frac{\alpha}{n} \sum (R_{opt} - R_i)^2$$

Der HOMA-Index basiert sowohl auf Bindungslängenalternanz als auch auf Bindungslängenvergrößerung und kann separiert werden in die Terme:

$$(1.35) \quad HOMA = 1 - EN - GEO$$

mit

$$(1.36) \quad EN = f \alpha (R_{opt} - R_{av})^2$$

$$(1.37) \quad f = \begin{cases} 1: R_{av} > R_{opt} \\ -1: R_{av} < R_{opt} \end{cases}$$

$$(1.38) \quad GEO = \frac{\alpha}{n} \sum_{i=1}^n (R_{av} - R_i)^2$$

Der Term EN steht hier für die Bindungslängenvergrößerung, GEO für die Bindungslängenalternanz.  $R_{av}$  steht für die mittlere Bindungslänge,  $R_i$  für den individuellen Bindungslängenabstand. Die HOMA-Methode wurde anhand von Röntgenstrukturdaten so parametrisiert, dass Benzol den hocharomatischen Wert 1 und 1,3-Butadien den nicht aromatischen Wert 0 erhalten.  $\alpha$  und  $R_{opt}$  sind dabei parametrisierte Konstanten siehe Tabelle 8



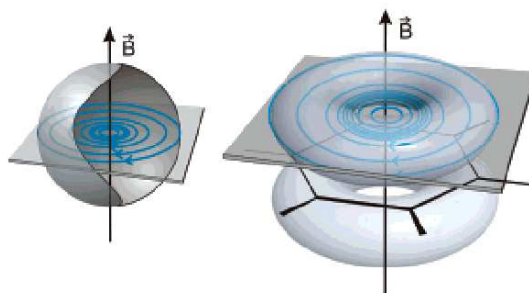
**Tabelle 8: Parametrisierte Konstanten der HOMA Methode**

bond	$R_{\text{opt}}$	$\alpha$
C-C	1.388	257.7

Weitere vom Programm *xyz2cane* ausgegebene Werte sind der maximale Bindungslängenabstand im betrachteten Ringsystem  $d_{\text{max}}$ , der minimale Bindungslängenabstand  $d_{\text{min}}$ , die Differenz dieser beiden Werte  $d_{\text{diff}}$ , sowie die maximale Abweichung aus der Planarität aller Diederwinkel im Raum in Grad,  $T_{\text{max}}$ .

#### 1.4.6 Anisotropie der induzierten Stromdichte (ACID)

Äußere Magnetfelder induzieren in Molekülen lokale und interatomare Ringströme, welche sich als „Abschirmung“ im NMR äußern. Diese kann je nach chemischer Umgebung charakteristische Werte annehmen. Lokale Ringströme (siehe Abbildung 16), die durch ein äußeres homogenes Magnetfeld induziert werden, sind um Größenordnungen stärker, als interatomare, wie z.B. der Ringstrom in Benzol. Im Gegensatz zu den interatomaren Ringströmen sind die lokalen Ringströme jedoch weitestgehend isotrop.



**Abbildung 16: Molekül im Magnetfeld, lokale (links) und interatomare (rechts) Ringströme**

Bei der ACID-Methode (*Anisotropy of the Induced Current Density*)<sup>[79],[80]</sup> wird das Skalarfeld des symmetrischen Anteils der Anisotropie der induzierten Stromdichte ACID-Funktion benannt und als Isooberfläche grafisch dargestellt (siehe Abbildung 17).

Sie kann als „Dichte delokalierter Elektronen“ sehr intuitiv verwendet werden. Im Falle konjugierter Bindungen und Hyperkonjugation ergeben sich genau dort große Werte, wo man Konjugation, respektive Hyperkonjugation erwartet. Im Falle aromatischer Moleküle kann sie überdies als „Aromatizität“ interpretiert werden, bei antiaromatischen Molekülen dementsprechend als Antiaromatizität.

Die ACID-Funktion erfüllt einige praktische Bedingungen:

- Sie ist ein Skalarfeld.

- Sie ist invariant gegenüber Ähnlichkeitstransformationen im Raum.
- Sie ist unabhängig gegenüber der relativen Orientierung des Moleküls zum äußeren Magnetfeld.
- Sie hat die gleiche Symmetrie wie die Gesamtwellenfunktion.
- Sie ist linear unabhängig von der Elektronendichtefunktion.

Mathematisch erhält man die ACID-Funktion aus dem Tensor der induzierten Stromdichte, welcher trotz großer Nachteile (insbesondere der Abhängigkeit von der relativen Orientierung des Moleküls zum äußeren Magnetfeld) oft zur Visualisierung aromatischer Ringströme verwendet wird. Die induzierte Stromdichte  $\vec{\mathbf{J}}^{(1)}$  ist gegeben durch:

$$(1.39) \quad \mathbf{T}^{(1)} = \begin{pmatrix} j_{xx} & j_{yx} & j_{zx} \\ j_{xy} & j_{yy} & j_{zy} \\ j_{xz} & j_{yz} & j_{zz} \end{pmatrix}$$

$$(1.40) \quad \vec{\mathbf{J}}^{(1)} = \mathbf{T}^{(1)} \cdot \vec{\mathbf{B}}$$

Die Anisotropie der induzierten Stromdichte  $\Delta\mathbf{T}^{(1)^2}$  wurde von Herges und Geuenich<sup>[79]</sup> definiert als Standardabweichung der Eigenwerte  $e_i$ .

$$(1.41) \quad \Delta\mathbf{T}^{(1)^2} = \sum_i (e_i - \bar{e})^2 \quad \text{mit}$$

$$\bar{e} = \frac{1}{n} \sum_i e_i \quad \text{und} \quad \sum_i e_i = \text{tr} \mathbf{T}$$

Die Zerlegung des Tensors (der Matrix)  $\Delta\mathbf{T}^{(1)}$  in einen symmetrischen und einen antisymmetrischen Anteil ergibt:

$$(1.42) \quad \Delta\mathbf{T}^{(1)^2} = \Delta\mathbf{T}_S^{(1)^2} + \text{tr} \mathbf{T}_A^{(1)} \mathbf{T}_A^{(1)\dagger}$$

Für die ACID-Funktion wird ausschließlich der symmetrische Anteil der Anisotropie  $\Delta\mathbf{T}_S^{(1)^2}$  verwendet. Analog zu Baders AIM-Methode (*Atoms in Molecules*)<sup>[93]</sup> kann ein kritischer *isosurface value* bestimmt werden, der sich dazu eignet, die AICD-Methode zu quantifizieren. Dabei ist der CIV (*critical isosurface value*) anschaulich als der Wert definiert, bei dem der Zusammenhalt der AICD-Oberfläche abreißt. Mathematisch handelt es sich um den Wert der ACID-Funktion an einem *critical point* der Isooberfläche. Dies ist ein Punkt, an dem die Ableitung des ACID-Skalarfelds nach allen Raumkoordinaten Null ist. Es gibt verschiedene Typen von kritischen Punkten, ausführlicher diskutiert in Baders AIM-Veröffentlichung.<sup>[93]</sup>

Als Beispiel einer Anwendung der Methode ist in Abbildung 17 der ACID-Plot des Farbstoffes Indigo gegenüber dem des Vorfarbstoffes Leukoindigo zu sehen. Wie deutlich zu erkennen ist, erstreckt sich die Konjugation im Indigo genau wie man es erwarten würde über beide Molekülhälften, während es beim farblosen Leukoindigo zu einer klaren Unterbrechung in der Mitte des Moleküls kommt.

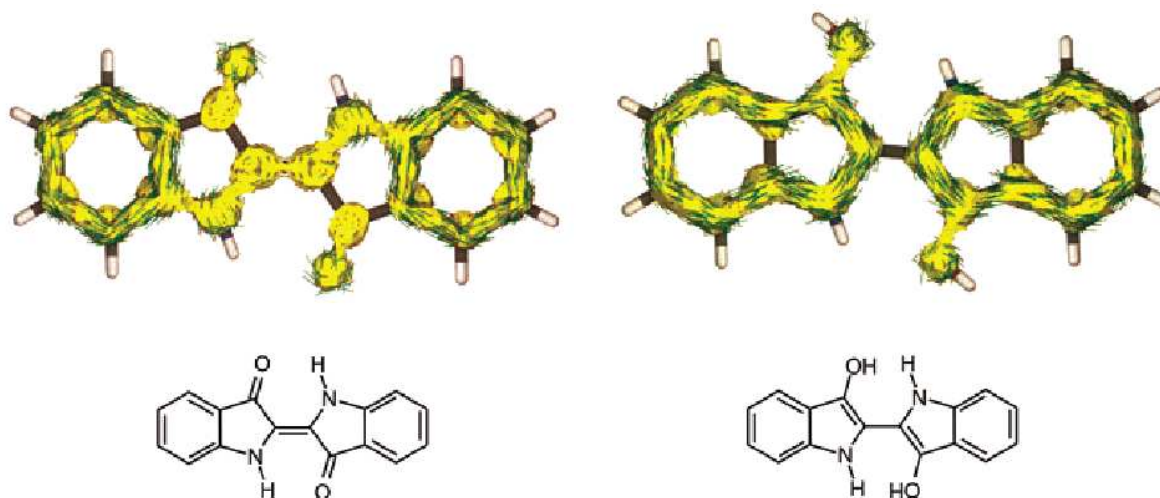


Abbildung 17: ACID Plot von Indigo (links) und Leukoindigo (rechts) entnommen aus<sup>[80]</sup>

Der ACID-Plot eignet sich somit hervorragend zur einfachen und intuitiven Visualisierung von Elektronendelokalisation auch in komplexen Molekülen, was im Annulenkaptel helfen sollte, die erhaltenen Ergebnisse visuell zu überprüfen.

## 1.5 Aufgabenstellung

### 1.5.1 Teil 1: Auswertung konformativer Untersuchungen mit Hilfe von Kohonenkarten

Der klassische Ansatz bei der Auswertung von MD-Rechnungen zur Konformationsanalyse besteht darin, in regelmäßigen Zeitabschnitten Strukturen zu minimieren. Es wird eine große Anzahl Minimumstrukturen erhalten, die jedoch häufig mehrfach auftauchen. Die Anzahl der Minimierungsoperationen wächst linear mit der simulierten Zeitspanne und ist der zeitbestimmende Faktor. Dies wäre aufgrund der Redundanz der Eingabedaten oft nicht nötig. Das Laufzeitverhalten jeder einzelnen Minimierung bewegt sich für Molekülmechanikrechnungen typischerweise im Bereich  $O(n^2)$  bis  $O(n^3)$ . Da eine große Anzahl Minimierungen nötig ist, erhält man selbst bei den einfachen physikalischen Modellen extrem lange Rechenzeiten.

Das strategische Ziel dieser Arbeit ist es, die Generalisierungsfähigkeit von neuronalen Netzen zu nutzen, um die Redundanz der Eingabedaten zu erniedrigen, Minima schon vor einer abschließenden Optimierung weitgehend zu charakterisieren sowie Beziehungen zwischen den Minima herauszufinden (Übergangszustände). Die Geometrien sollen topologisch richtig auf eine zwei-dimensionale Karte abgebildet werden, um die Ergebnisse der MD-Simulation dem menschlichen Erfassungsvermögen zugänglich zu machen. Mathematisch ausgedrückt bedeutet dies Folgendes: Es soll eine Projektion einer hochdimensionalen Eingabedatenmenge topologieerhaltend auf die Ebene projiziert werden.

Dafür gehen wir von folgender Voraussetzung aus: Wir verwenden einen Satz Eingabedaten, in dem alle Molekülgeometrien von Interesse vorhanden sind. Üblicherweise ist dies z.B. eine Moleküldynamikrechnung, eine Monte-Carlo-Simulation oder ein Satz generischer Daten (mittels heuristischer Methoden gezielt erzeugte Geometrien). Diese Daten werden aufbereitet und mit ihnen ein im Rahmen dieser Arbeit erstelltes Kohonennetzprogramm trainiert.

Das Netzwerk soll nun Folgendes leisten: Es soll die Konformationen aus den durch thermische Schwingungen der Atome verrauschten Eingabedaten extrahieren und auf einer 2-dimensionalen Karte abbilden. Durch Abfrage bestimmter Molekülparameter wie z.B. Bindungslängen sollen in  $O(1)$ -Laufzeitverhalten (bezogen auf die Länge der Simulation und die Größe der Moleküle) passende Geometrien gefunden werden, was Perspektiven im *rational drug design* eröffnen dürfte. Ein Hauptziel ist des Weiteren eine interaktive Visualisierung der Trajektorie des Systems (falls anwendbar, also bei MD- und MC-Auswertungen).

### **1.5.2 Teil 2: Aromatizität von $[4n]$ Annulen-Isomeren**

Nach der ersten Synthese und eindeutigen Strukturcharakterisierung durch Röntgenstrukturanalyse durch Herges und Ajami<sup>[84]</sup> richtet sich das Augenmerk in der Synthesechemie erneut auf das Prinzip der Möbiusaromatizität.

Bereits in den Stammsystemen der ungeladenen zyklischen Annulene mit  $4n$   $\pi$ -Elektronen besteht eventuell die Möglichkeit, dass möbiusartige Strukturen eine signifikante Rolle im Isomerenensemble einnehmen. Diese Hypothese soll für Annulene unterschiedlicher Größe untersucht werden.

Geladene  $[4n]$ -Annulene sind bisher hauptsächlich rechnerischen Verfahren zugänglich. Im Rahmen dieser Arbeit sollen neue Methoden zur effizienten Erzeugung und Auswertung einer großen Anzahl von Annulen-Isomeren

entwickelt werden. Dazu sollen das Programm Gaussian03<sup>[2]</sup> sowie selbst implementierte Programme und Verfahren Anwendung finden. Diese Arbeit hofft einen wesentlichen Beitrag zur Klärung der Frage zu liefern, ob Möbiusaromatizität in diesen Molekülen tatsächlich vorliegt und wie realistisch eine synthetische Darstellung von geladenen möbiusartigen Annulenen sein wird.

Gewissermaßen ein Nebenprodukt ist dabei die Implementierung und Erprobung der FISE-Methode, welche schwache aromatische Einflüsse in gespannten nichtplanaren Ringen besser quantifizierbar machen soll.

## 2 Kohonnennetze zur Konformationsanalyse

Bei den Kohonnennetzen (*Kohonen feature maps*) handelt es sich um selbstorganisierende künstliche Netzwerke, zuerst vorgeschlagen von Teuvo Kohonen.<sup>[94]</sup> Kohonnennetze dienen zur topologieerhaltenden Projektion eines hochdimensionalen Eingangsraums in einen niedrigdimensionalen Zielraum.

$$(2.1) \quad \mathbf{X} = \begin{pmatrix} x_1 \\ \vdots \\ x_M \end{pmatrix} \rightarrow \mathbf{Y} = \begin{pmatrix} y_1 \\ \vdots \\ y_N \end{pmatrix}$$

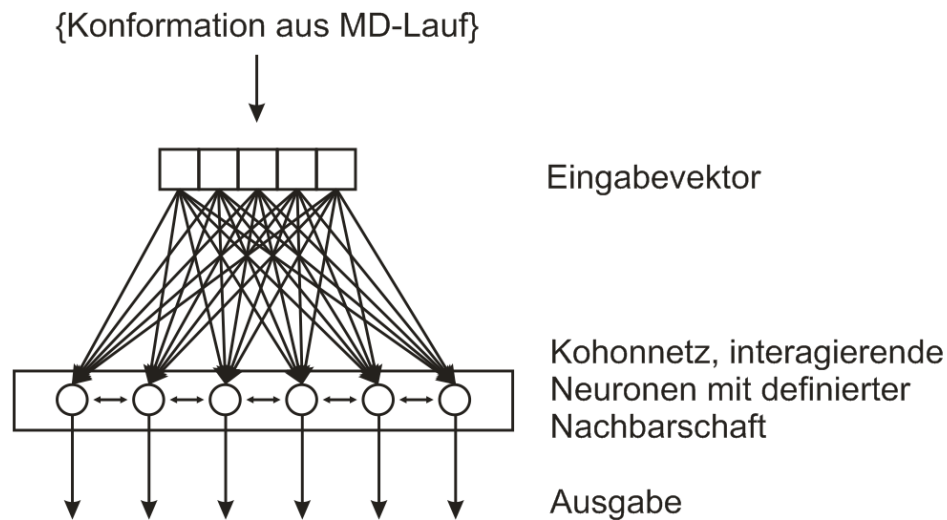
Es sei  $M$  die Dimensionalität des Ausgangsraums,  $N$  die Dimensionalität des Zielraums und  $M \geq N$ . Topologieerhaltung bedeutet, dass Nachbarschaftsbeziehungen, definiert als niedriger Abstandswert in Gleichung (1.12), die im ursprünglichen Koordinatenraum bestehen, möglichst auch im Zielraum erhalten bleiben sollen. Inzwischen gibt es zahllose Varianten, die sich meist nur in Implementationsdetails unterscheiden. Allen gemeinsam sind als wichtigste Eigenschaften das unbeaufsichtigte Lernen sowie die Erhaltung der Nachbarschaftsbeziehungen bei der Projektion. Die bei weitem gebräuchlichste Projektion ist die von Kohonen<sup>[94]</sup> vorgeschlagene Projektion in den zweidimensionalen Raum, also  $N=2$ . Nicht jeder Datensatz eignet sich für eine Projektion, es kann zwangsläufig zu Projektionsfehlern kommen, an denen Nachbarschaftsbeziehungen „auseinander gerissen“ werden oder ursprünglich nicht benachbarte Datensätze im Zielraum benachbart sind. In diesen Fällen wurden von Pawelzik/Bauer höherdimensionale Kohonenkarten vorgeschlagen, wobei das „Waber Produkt“ eine Funktion zur Ermittlung der geeigneten Dimensionierung darstellen soll.<sup>[95],[96]</sup>

Damit verlieren die projizierten Daten aber ihre Übersichtlichkeit und damit ihren größten Nutzen, weswegen sich diese Arbeit auf zweidimensionale Karten beschränken wird.

### 2.1 Topologie der Kohonenkarte

Kohonnennetze sind wahrscheinlich von allen Netzwerkarten in Architektur und Lernverhalten am ehesten mit biologischen Systemen vergleichbar. Insbesondere der visuelle Kortex in Gehirnen höher entwickelter Lebewesen ist bereits zum Ziel direkter experimenteller Vergleiche mit diesen künstlichen neuronalen Netzen geworden.<sup>[97]</sup>

Nach der Nomenklatur aus Kapitel 1.3.5.2 handelt es sich um ein Einlagen-Design wie in Abbildung 9 dargestellt, in dem die Neuronen in einer Ebene angeordnet sind, auch *competitive layer* genannt. Als zusätzliches Element sind die Neuronen mit ihren Nachbarneuronen verknüpft und beeinflussen diese während des Lernvorgangs.

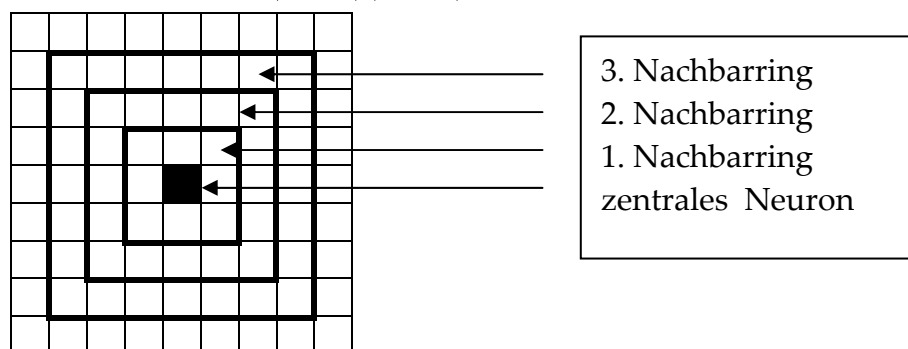


**Abbildung 18:** Schema der Architektur eines Kohonennetzes, in Abwandlung des allgemeinen Netzschemas in Abbildung 9 mit zusätzlichen lateralen Verknüpfungen

Aus den Nachbarschaftsbeziehungen erhält das Netz eine genau definierte Topologie. In dieser Arbeit wird eine topologisch quadratische Nachbarschaft definiert. In Anlehnung an das Schachspiel ist die Distanz zweier Neuronen auf der Karte die minimale Anzahl der Königszüge, um vom einen zum anderen zu gelangen.

(2.2)

$$\Delta = \max(|x - x_c|, |y - y_c|)$$



**Abbildung 19:** Die quadratische topologische Nachbarschaft eines Neurons

Denkbar wäre es auch, die Nachbarschaft hexagonal oder als euklidische Distanz (rationale Zahl) auf der Netzoberfläche zu definieren. Eine weitere oft verwendete Abstandsgleichung ist die  $L_1$ -Norm (*Manhattan distance*)

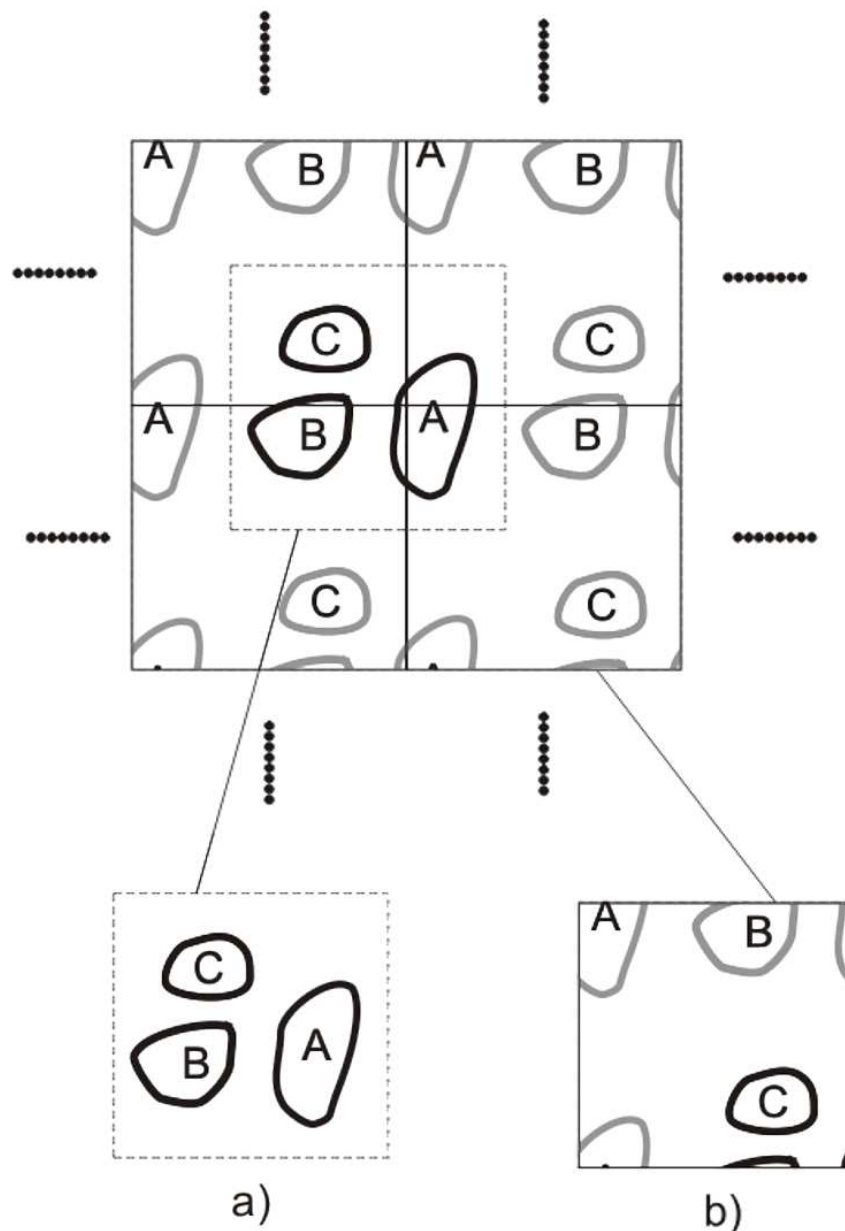
$\Delta = |x - x_c| + |y - y_c|$ , welche der Anzahl der Turmzüge der Länge 1 entspricht, welche nötig sind um von einem Feld zum anderen zu gelangen.

## 2.2 Grafische Repräsentation der Kohonenkarte

Bei den Kohonenkarten handelt es sich topologisch gesehen um die Oberfläche eines Torus, aufgeschnitten und projiziert in die Ebene. Aus Anschaulichkeitsgründen wird die Netzkarte im Rahmen dieser Doktorarbeit immer als Quadrat präsentiert. Aus der Torustopologie folgt die Abwesenheit von Rändern. Ein Überschreiten des oberen Randes bedeutet demnach, dass man sich am unteren Rand wieder findet; entsprechend sind auch linke und rechte Ränder miteinander verkettet.

Um das Grenzproblem zu umgehen, können vier identische Karten in Rechteckform aneinander gehängt werden, so dass jedes Neuron stets mindestens einmal mit seiner Nachbarschaft in allen Richtungen zu sehen ist. Will man jedes Neuron nur einmal abbilden, führt dies zum Problem, dass je nach Wahl des betrachteten Fensters die Oberfläche sehr unterschiedlich aussehen kann.





**Abbildung 20:** Oberfläche eines Kohonenetzes, a) geschickte und b) ungeschickte Wahl des Projektionsausschnitts

Generell ist es sinnvoll, die zu betrachtenden Konformationsbereiche möglichst in das Zentrum des betrachteten Fensters zu legen. Exemplarisch wird dies in Abbildung 20 anhand einer „guten“ und einer „schlechten“ Wahl des Projektionsfensters dargestellt. Die Willkürlichkeit der Festlegung des Schnittes sollte man sich immer vor Augen halten.

Die Oberfläche kann eingefärbt werden mit Eigenschaften des ursprünglichen Datenraumes  $\mathbf{X}$  aus der Projektion in Gleichung (2.1).

### 2.3 Training von Kohonennetzen – allgemeiner Teil

Ein Kohonennetz wird durch unbeaufsichtigtes Lernen<sup>[98]</sup> in einem iterativen Prozess trainiert (*unsupervised learning*, auch *self-learning*). Das bedeutet, dass es kein Feedback für das Netz gibt, wie gut oder schlecht seine derzeitigen Gewichte gewählt sind. Der Input in ein Kohonennetz besteht aus einer endlichen Menge an Trainingsstimuli  $\mathbf{A}$ , denen jeweils ein Eingabevektor  $\mathbf{X}_i$  zugeordnet wird

$$(2.3) \quad \mathbf{A} = \{A_i = (\mathbf{X}_i) \mid i = 1, \dots, \mu_a\} \quad \mu_a : \text{Anzahl der Stimuli}$$

Der Eingabevektor  $\mathbf{X}_i$  ist in Anlehnung an die Projektionsgleichung(2.1) ein M-dimensionaler Vektor, wobei M die Dimensionalität des ursprünglichen Parameterraums angibt. Üblicherweise und auch im Rahmen dieser Arbeit werden die Stimuli in normalisierter Form verwendet. Ein vollständiger Durchgang aller Stimuli  $\mathbf{A}$  wird dabei als *epoch* bezeichnet.

$$(2.4) \quad \mathbf{X}_i = \begin{pmatrix} x_{i,1} \\ \vdots \\ x_{i,p} \end{pmatrix} \quad i = 1 \dots \mu_a, p = 1 \dots M \quad 0 \leq x_{i,p} \leq 1$$

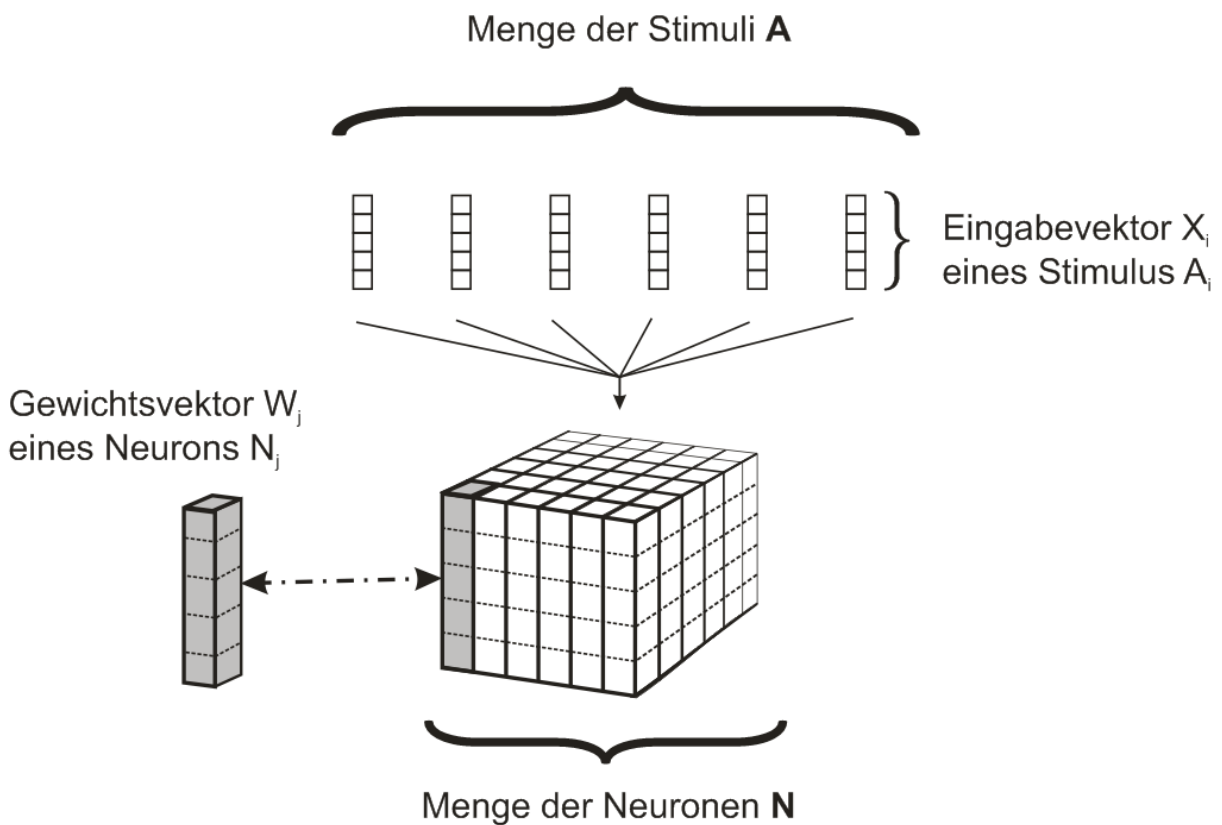
Die Datenstruktur eines Kohonennetzes besteht aus einer Menge Neuronen  $\mathbf{N}$ . Jedem Neuron ist ein Gewichtsvektor  $\mathbf{W}_j$  zugeordnet.

$$(2.5) \quad \mathbf{N} = \{N_j = (\mathbf{W}_j) \mid j = 1, \dots, \mu_n\} \quad \mu_n : \text{Anzahl der Neuronen}$$

Auch der Gewichtsvektor ist M-dimensional und normalisiert.

$$(2.6) \quad \mathbf{W}_j = \begin{pmatrix} w_{j,1} \\ \vdots \\ w_{j,p} \end{pmatrix} \quad j = 1 \dots \mu_n, p = 1 \dots M \quad 0 \leq w_{j,p} \leq 1$$

Dargestellt werden diese Definitionen in Abbildung 21.



**Abbildung 21: Datenstrukturen einer Kohonenkarte**

Die metrische Distanz  $d(A_i, N_j)$  eines Stimulus zu einem Neuron ist definiert als die euklidische Distanz des zugeordneten Eingabevektors mit dem Gewichtsvektor des Neurons.

$$(2.7) \quad d(A_i, N_j) \equiv d(\mathbf{X}_i, \mathbf{W}_j)$$

$$d(\mathbf{X}_i, \mathbf{W}_j)^2 = \sum_{p=1}^M d_{euclid}(X_{i,p}, W_{j,p})^2$$

Analog dazu ist auch die metrische Distanz zwischen zwei Neuronen  $N_j$  und  $N_k$  definiert in Gleichung (2.8) sowie die zweier Stimuli  $A_j$  und  $A_k$  durch Gleichung (2.9).

$$(2.8) \quad d(\mathbf{W}_j, \mathbf{W}_k)^2 = \sum_{p=1}^M d_{euclid}(W_{j,p}, W_{k,p})^2$$

$$(2.9) \quad d(\mathbf{X}_j, \mathbf{X}_k)^2 = \sum_{p=1}^M d_{euclid}(X_{j,p}, X_{k,p})^2$$

Die Verwendung der metrischen Distanz aus Gleichung (2.7) zwischen Stimulus und Neuron als Unähnlichkeitskriterium wurde bereits in Kapitel 1.3 vorgestellt.

In der Trainingsphase werden die Gewichte des Netzes zunächst zufällig und gleichverteilt initialisiert. In der Folge werden dem Netz alle Stimuli in zufäl-

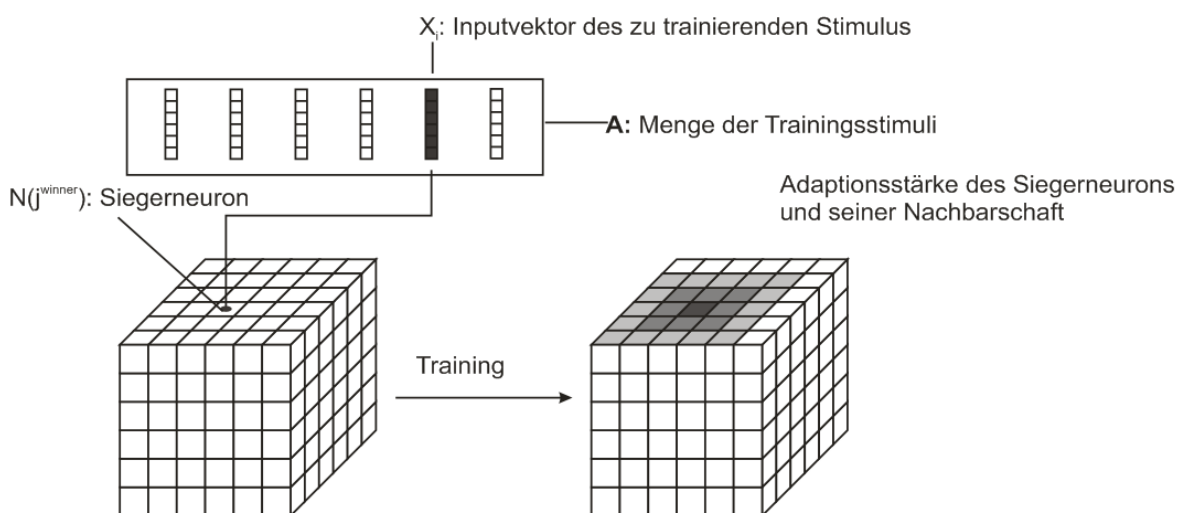
liger Reihenfolge präsentiert. Zu jedem Stimulus  $A_i$  wird ein Erregungszentrum gefunden, welches das Neuron mit der niedrigsten metrischen Distanz  $d(A_i, N_j)$  ist. Da in diesem Schritt die Neuronen quasi um die Stimuli konkurrieren, wird das Verfahren auch *competitive learning* genannt.

$$(2.10) \quad j^{winner} \leftarrow \min(d(A_i, N_j)) \quad \forall j = 1 \dots \mu_n$$

In Analogie zur Hebbischen Lernregel<sup>[99]</sup> („*what fires together, wires together*“) werden jetzt die Gewichte des Siegerneurons (im Modell von Abbildung 7 stehen diese direkt für die Axonstärken) und seiner Nachbarschaft dem Stimulus weiter angepasst.

$$(2.11) \quad \mathbf{W}_j^{(neu)} = \mathbf{W}_j^{(alt)} + f \cdot (\mathbf{X}_i - \mathbf{W}_j)$$

Eine grafische Repräsentation eines Trainingsschrittes soll den Prozess verdeutlichen. (siehe Abbildung 22)



**Abbildung 22: Grafische Darstellung eines Trainingsschrittes**

Die Anpassungsstärke  $f$  aus Gleichung (2.11) wurde von Kohonen<sup>[94]</sup> Adaptionfaktor bezeichnet (auch Adaptionstärke bei anderen Autoren).

$$(2.12) \quad f = f_0 \cdot \eta(\xi) \cdot \vartheta(\zeta)$$

$\eta(\xi)$ : Zeitabhängigkeit des Adaptionfaktors

$\vartheta(\zeta)$ : Abstandsabhängigkeit des Adaptionfaktors (auch Nachbarschaftsfunktion oder „*neighborhood function*“)

$f_0$ : maximaler Adaptionfaktor

In dieser Arbeit wird analog zur Reaktionslaufzahl in der chemischen Kinetik eine Trainingslaufzahl  $\xi$  definiert. Sie bestimmt die Stärke des Adaption-

schritts in Abhängigkeit von Reaktionsfortschritt und Entfernung vom Erregungszentrum. In dieser Arbeit ist sie definiert gemäß Gleichung (2.13).

$$(2.13) \quad \xi = \frac{t}{t_{ges}}$$

t: Zahl der schon trainierten Stimuli

$t_{ges}$ : Gesamtzahl der zu trainierenden Stimuli

Des Weiteren wird eine relative topologische Entfernungsfunktion definiert, die sich aus der topologischen Distanz  $\Delta$  errechnet nach

$$(2.14) \quad \zeta = \frac{\Delta}{\Delta_{max}(\xi)}$$

$\Delta$ : topologische Distanz zwischen Neuron und Erregungszentrum nach Gleichung (2.2)

$\Delta_{max}(\xi)$ : maximale Trainingsdistanz, verringert sich mit zunehmendem Trainingsfortschritt.

Für die maximale Trainingsdistanz wird in dieser Arbeit eine linear fallende Funktion des Trainingsfortschritts verwendet.

$$(2.15) \quad \Delta_{max}(\xi) = \begin{cases} \Delta_{var} = \Delta_0 \cdot \omega \cdot \xi + 1 & \forall \Delta_{var} < \omega \\ \omega & \forall \Delta_{var} \geq \omega \end{cases}$$

$\Delta_0$ : Konnektivitätsparameter,  $0 \leq \Delta_0 \leq 1.0$

$\omega$ : maximaler Lernradius

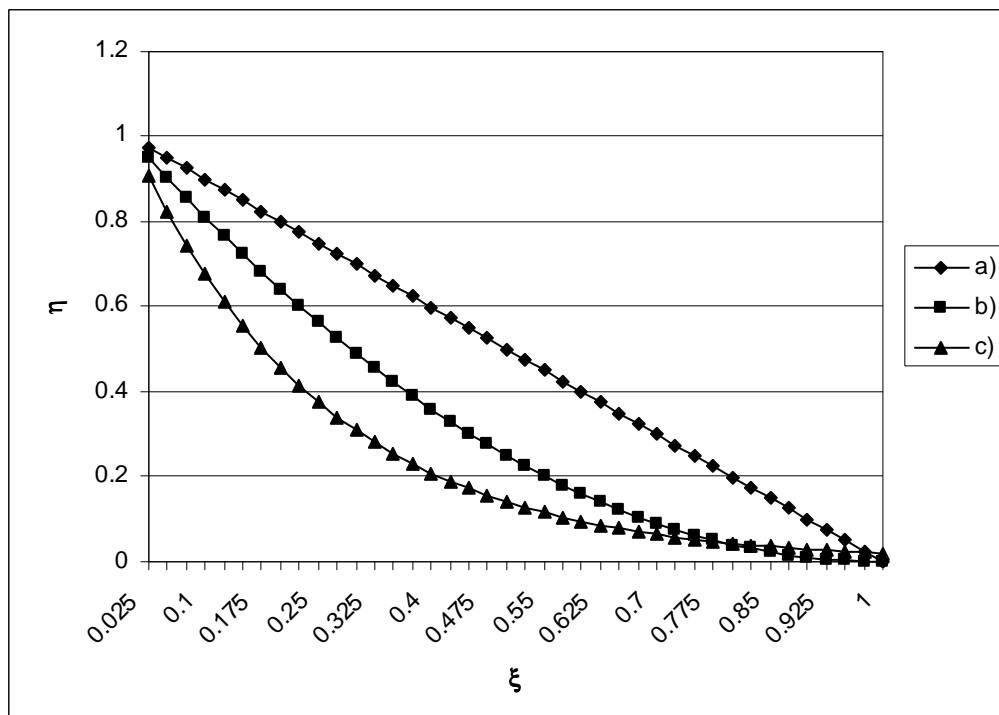
$$(2.16) \quad \omega = \min\left(\frac{\dim(\mathbf{N}, x) - 1}{2}, \frac{\dim(\mathbf{N}, y) - 1}{2}\right)$$

Der Konnektivitätsparameter  $\Delta_0$  gibt an, wie weit reichend die lateralen Verknüpfungen der Neuronen untereinander sind. Es hat sich bei Versuchen in dieser Arbeit gezeigt, dass eine wesentliche Beschleunigung des Trainings erreicht werden kann, wenn von Anfang an nicht das gesamte Netz als Nachbarschaft des Siegerneurons mittrainiert wird. Wählt man den Wert zu klein, werden allerdings Stimuli aus  $\mathbf{A}$  mehrfach auf der Netzkarte abgebildet, was auch bei vollständiger Abdeckung der Nachbarschaft vorkommen kann, aber unwahrscheinlicher ist. Aus der Periodizität der Netztopologie ergibt sich ein maximaler Lernradius (siehe Gleichung (2.16)), wenn vermieden werden soll, dass ein Neuron mehrfach mittrainiert wird.

Für die Zeitabhängigkeit des Adaptionfaktors  $\eta(\xi)$  werden verschiedene Funktionen implementiert, siehe (2.17).

$$(2.17) \quad \begin{aligned} a) \quad & \eta(\xi) = (1 - \xi) \\ b) \quad & \eta(\xi) = (1 - \xi)^2 \\ c) \quad & \eta(\xi) = \eta_{end}^\xi \end{aligned}$$

Die besten Trainingsraten werden mit den Varianten a) und c) erzielt, was sich mit Literaturergebnissen deckt.<sup>[100]</sup> Der Einfluss ist jedoch, wie sich in den folgenden Kapiteln zeigen wird, äußerst gering. Das wesentliche Kriterium scheint zu sein, dass die Funktion monoton fallend ist. Wenn nicht anders angegeben, wird deshalb die einfachste Variante a) der Zeitabhängigkeit aus Gleichung (2.17) verwendet. Ritter *et al.* bevorzugen die Exponentialfunktion c), welche eine Art *Simulated Annealing* der Gewichte der Neuronen beinhaltet.



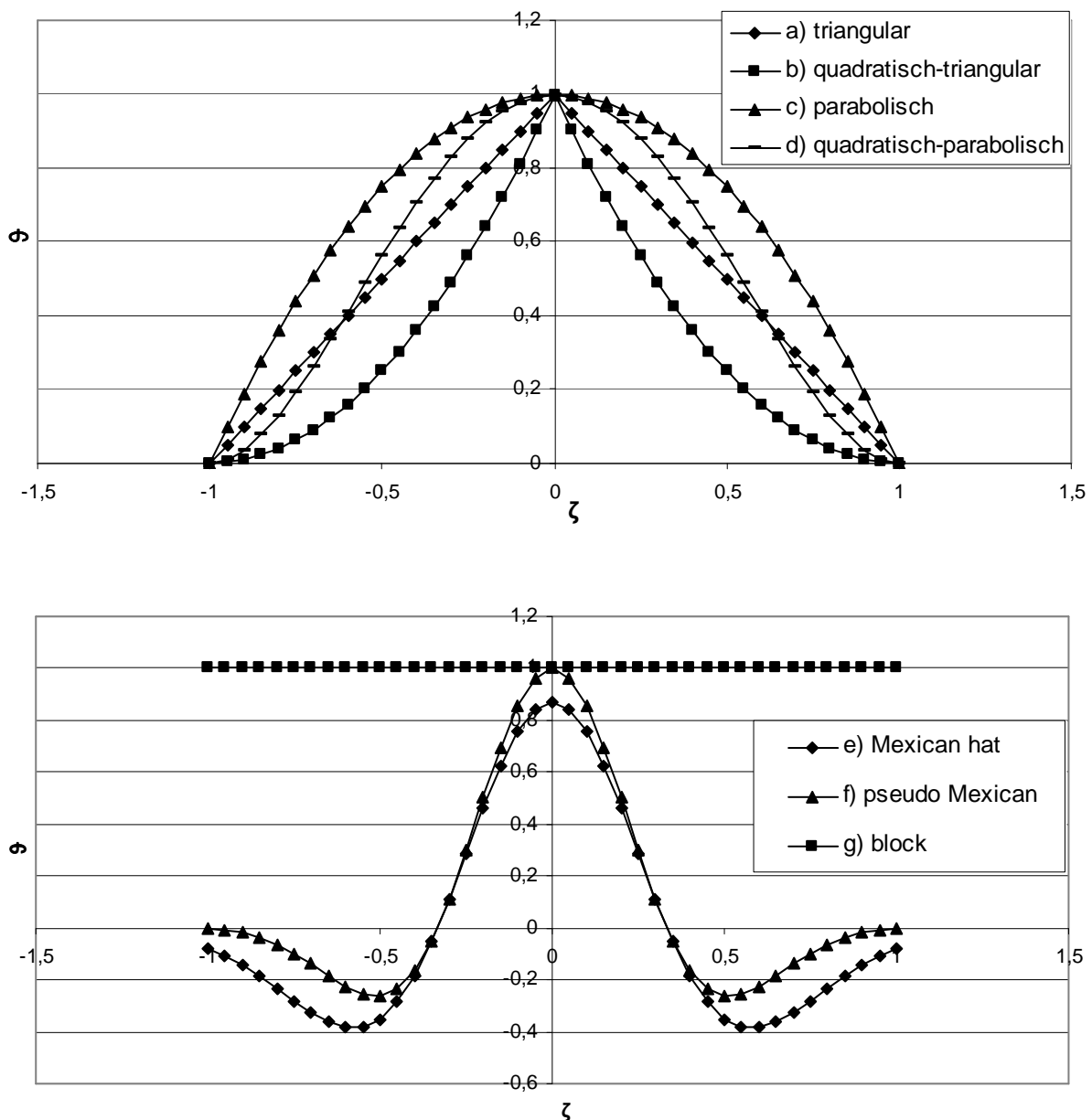
**Abbildung 23:** Zeitabhängigkeit des Adaptionfaktors  $f$ , verschiedene Funktionen

Auch für die Abstandabhängigkeit  $\vartheta(\Delta, \xi)$  können wieder verschiedene Funktionsformen herangezogen werden. Folgende werden im Rahmen dieser Arbeit implementiert.

(2.18)

- |    |  |                                  |
|----|--|----------------------------------|
| a) | $\vartheta(\zeta) = 1 -  \zeta $   | <i>triangular</i>                |
| b) | $\vartheta(\zeta) = (1 -  \zeta )^2$   | <i>quadratisch – triangular</i>  |
| c) | $\vartheta(\zeta) = (1 - \zeta^2)$   | <i>parabolisch</i>               |
| d) | $\vartheta(\zeta) = (1 - \zeta^2)^2$   | <i>quadratisch – parabolisch</i> |
| e) | $\vartheta(\zeta) = \left(\frac{2}{\sqrt{3}} \pi^{-1/4}\right) \left(1 - \left(\frac{\zeta}{3}\right)^2\right) \exp(-(\zeta/3)^2/2)$ | <i>Mexican hat</i>               |
| f) | $\vartheta(\zeta) = \cos\left(\frac{3}{2}\zeta\right) \exp(-4\zeta^2)$   | <i>pseudo Mexican</i>            |
| g) | $\vartheta(\zeta) = 1$   | <i>block</i>                     |

Grafisch aufgetragen in Abbildung 24 ist zu erkennen, dass es zwei Klassen von Nachbarschaftsfunktionen gibt, solche mit einem Nulldurchgang und solche ohne. Die negative Lernrate der entfernten Neuronen in der „Mexican hat“ Funktion und der in dieser Arbeit mit „pseudo Mexican“ benannten Funktionen hat dabei das Vorbild in der Natur in Form der „lateralen Hemmung“. Neuronen werden durch aktive Nachbarneuronen im Lernen gehemmt.



**Abbildung 24: Abstandsabhängigkeit des Adaptionfaktors  $f$ , verschiedene Funktionen**

Die einfache triangulare Variante a) wurde schon mit guten Ergebnissen in Kohonens Originalpublikation<sup>[101]</sup> verwendet. Das besondere Feature der sehr bekannten „Mexican hat“ Funktion liegt in der ausgezeichneten Trennung der auf die Netzkarte projizierten Cluster durch die negative Lernrate am Rand. Für Clusteringanwendungen ist sie oft die erste Wahl. Problematisch ist in diesem Zusammenhang die Tatsache, dass die Grenzflächen zwischen den Clustern mitunter nicht als solche erkannt werden können. In dem Fall erscheinen diese Neuronen als künstliche Artefakte und bewirken, dass scheinbar Daten auf der Netzkarte abgebildet werden, die nicht im ursprünglichen Datensatz vorkommen.



Die in dieser Arbeit vorgeschlagene „*pseudo Mexican*“ Funktion e) erfüllt ähnliche Zwecke wie die „*Mexican hat*“ Funktion.  $\vartheta(\zeta=\pm 1)$  ist jedoch Null, damit erhält man stetiges Verhalten am Rande, was bei der Implementierung Vorteile mit sich bringt.

## 2.4 Training von Kohonennetzen zur Konformationsanalyse

Wie in Kapitel 1.5.1 beschrieben, besteht der erste Schritt der Konformationsanalyse in der Erstellung eines Satzes an Stimuli  $\mathbf{A}$ . Aus diesen werden dann die Eingabevektoren für das Training extrahiert. Formal ausgedrückt, handelt es sich um die Projektion:

$$(2.19) \quad \mathbf{A} = \begin{pmatrix} k_1 \\ \vdots \\ k_L \end{pmatrix} \rightarrow \mathbf{X} = \begin{pmatrix} x_1 \\ \vdots \\ x_M \end{pmatrix}$$

Im Rahmen dieser Arbeit wird ein Konformerensemble im dreidimensionalen kartesischen Koordinatenraum als Satz  $\mathbf{A}$  verwendet, da kartesische Koordinaten direkt aus allen vorgeschalteten Programmen und Methoden erhältlich sind. Ein N-atomiges Molekül wird mit  $L=3N$  Dimensionen im kartesischen Raum beschrieben (siehe Gleichung (1.1), hat aber nur  $3N-6$  interne Freiheitsgrade (ohne Translation und Rotation).

Die Darstellung in kartesischen Koordinaten ist nicht invariant gegenüber Translationen und Rotationen. Dies erschwert den Vergleich von Molekülkonformationen untereinander. Aus kartesischen Koordinaten lassen sich interne Koordinaten gewinnen, welche invariant gegenüber Rotation und Translation des Koordinatenraums sind. In dieser Arbeit werden drei Arten von internen kartesischen Koordinaten verwendet: interatomare Abstände, Winkel und Torsionswinkel. Die Auswahl der verwendeten Koordinaten geschieht manuell. Genau wie in anderen Methoden (siehe Kapitel 1.3) ist die Wahl der in  $\mathbf{X}$  abzubildenden internen Koordinaten ein Schlüsselschritt. In der Regel werden deutlich weniger Koordinaten trainiert als das Molekül Freiheitsgrade hat. Werden hier zu wenige oder falsche Koordinaten gewählt, ist die Molekülstruktur nicht ausreichend in  $\mathbf{X}$  kodiert. Stark unterschiedliche Moleküle in  $\mathbf{K}$  werden somit in  $\mathbf{X}$  ununterscheidbar, die erhaltenen Netzkarten sind unbrauchbar zur Beschreibung des Konformationsraums. Werden zu viele Koordinaten definiert, z.B. die vollständige Abstandsmatrix aller Atome miteinander, wächst der Parameterraum und damit die Rechenzeit quadratisch. Außerdem sind die erhaltenen Netzkarten oft wenig aussagekräftig, Beispiele dazu werden in nachfolgenden Kapiteln gezeigt werden. Eindeutige

und objektive Kriterien für eine „gute“ Wahl der Parameter konnten in dieser Arbeit nicht gefunden werden. Damit ist das Verfahren, genau wie viele andere, von der Erfahrung und dem „Fingerspitzengefühl“ des Anwenders abhängig.

Folgende zwei Regeln zur Auswahl „guter“ interner Koordinaten werden hiermit als „Struktur/Funktions-Prinzip“ postuliert:

1. Die Struktur des Moleküls muss durch die zu trainierenden internen Koordinaten möglichst eindeutig definiert sein.
2. Die für den Anwender interessante Eigenschaft oder Funktion des Moleküls sollte durch die zu trainierenden Koordinaten beschrieben werden.

Damit trainiert man meist einen Satz aus zwei Gruppen von internen Koordinaten:

1. Torsionswinkel eignen sich besonders zur Strukturbeschreibung, analog zur systematischen Konformationsuche, siehe dazu auch die Arbeit von Guida, Saunders *et al.*<sup>[27]</sup>
2. Abstände geben oft kritische Eigenschaften für das Verhalten eines Moleküls wieder, z.B. ob bei Dockingstudien eine Bindung eines Substrats erfolgt, oder ob eine offene oder geschlossene Struktur vorliegt.

Die Transformation der kartesischen Koordinaten in interne Koordinaten samt Normalisierung in Gleichung (2.19) erfolgt dabei durch folgende Vorschriften:

(2.20)

$$\begin{array}{ll}
 a) & x_{i,p}^{(Dist)} = \frac{koord_{i,p}^{(Dist)} - koord_{p,min}^{(Dist)}}{koord_{p,max}^{(Dist)} - koord_{p,min}^{(Dist)}} \quad \text{Abstände (Distanzen)} \\
 b) & x_{i,p}^{(Winkel)} = \frac{koord_{i,p}^{(Winkel)} - koord_{p,min}^{(Winkel)}}{koord_{p,max}^{(Winkel)} - koord_{p,min}^{(Winkel)}} \quad \text{Winkel} \\
 c) & x_{i,p}^{(Tors)} = \frac{koord_{i,p}^{(Tors)}}{2\pi} \quad \text{Torsionswinkel}
 \end{array}$$

Die erhaltenen  $X_i$  entsprechen den Stimuli wie in Gleichung (2.4) angegeben. Die Eingabewerte werden innerhalb jeder Dimension des Stimulus separat normalisiert. Dafür werden im Rahmen des Trainingssatzes der größte für eine interne Koordinate vorkommende Wert  $koord_{p,max}$  und der kleinste vorkommende Wert  $koord_{p,min}$  herangezogen. Wie bereits in Gleichung (2.20) zu sehen, stellen Torsionswinkel einen speziellen Fall dar. Dies liegt darin begründet, dass Torsionswinkel einen zyklischen Wertebereich besitzen. Dies

hat Auswirkungen auf die zu verwendende Distanzfunktion. Im Rahmen dieser Arbeit wird die euklidische Distanz zweier Werte im Raum der Torsionswinkel definiert durch Gleichung (2.21), bzw. (2.22).

$$(2.21) \quad d_{euclid}^{Tors} (koord_1^{Tors}, koord_2^{Tors}) = \min \left( \begin{array}{l} |koord_1^{Tors} - koord_2^{Tors}|, \\ 2\pi - |koord_1^{Tors} - koord_2^{Tors}| \end{array} \right),$$

nach Normalisierung der Koordinaten erhält man:

$$(2.22) \quad d_{euclid}^{Tors} (X_{i,p}, W_{j,p}) = \min \left( \begin{array}{l} |X_{i,p} - W_{j,p}|, \\ 1 - |X_{i,p} - W_{j,p}| \end{array} \right)$$

Mit dieser Definition (Gleichung (2.22)) kann Gleichung (2.7) verwendet werden, unabhängig von der Art der trainierten internen Koordinaten (Abstände, Winkel, Diederwinkel).

Schematisch beschrieben wird das Training von den Konformeren zu den Eingabevektoren bis hin zum fertig trainierten Netz in Abbildung 25.

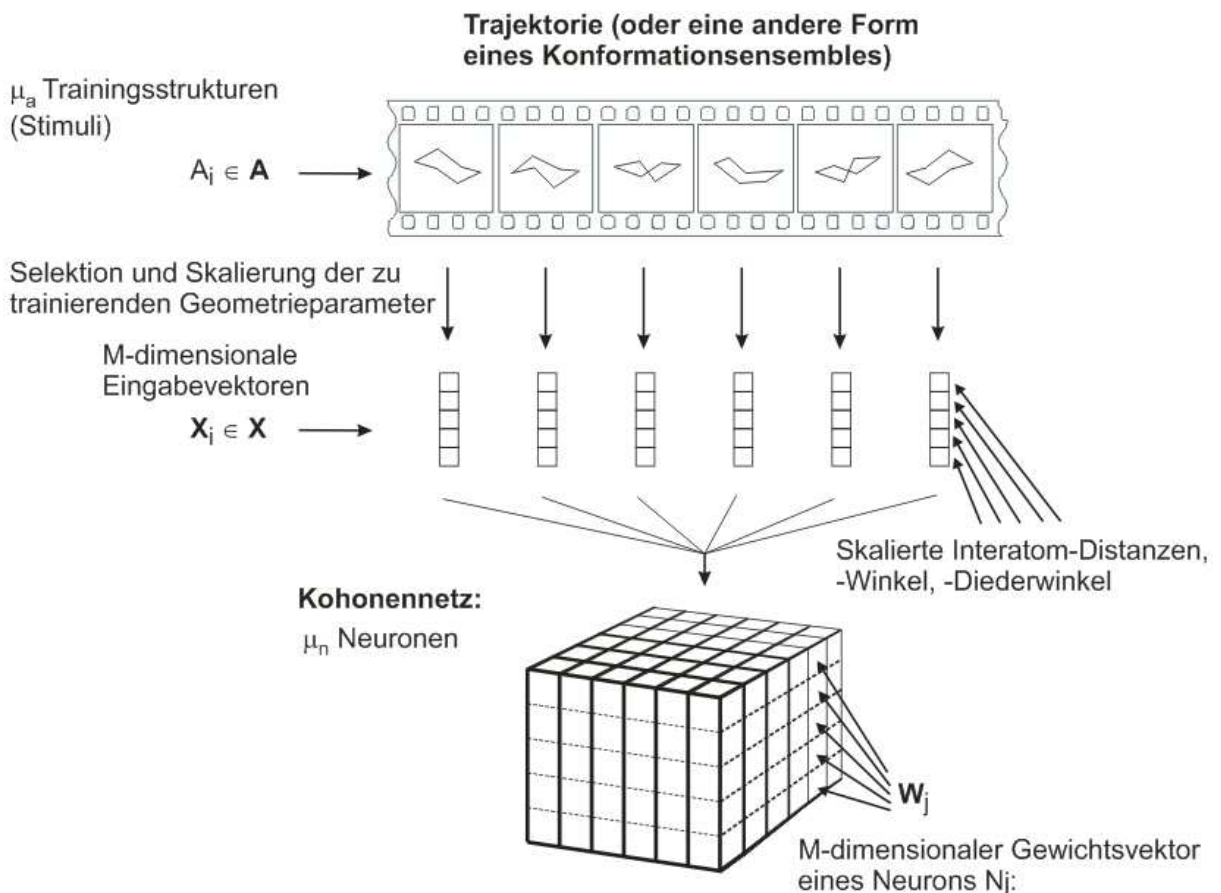
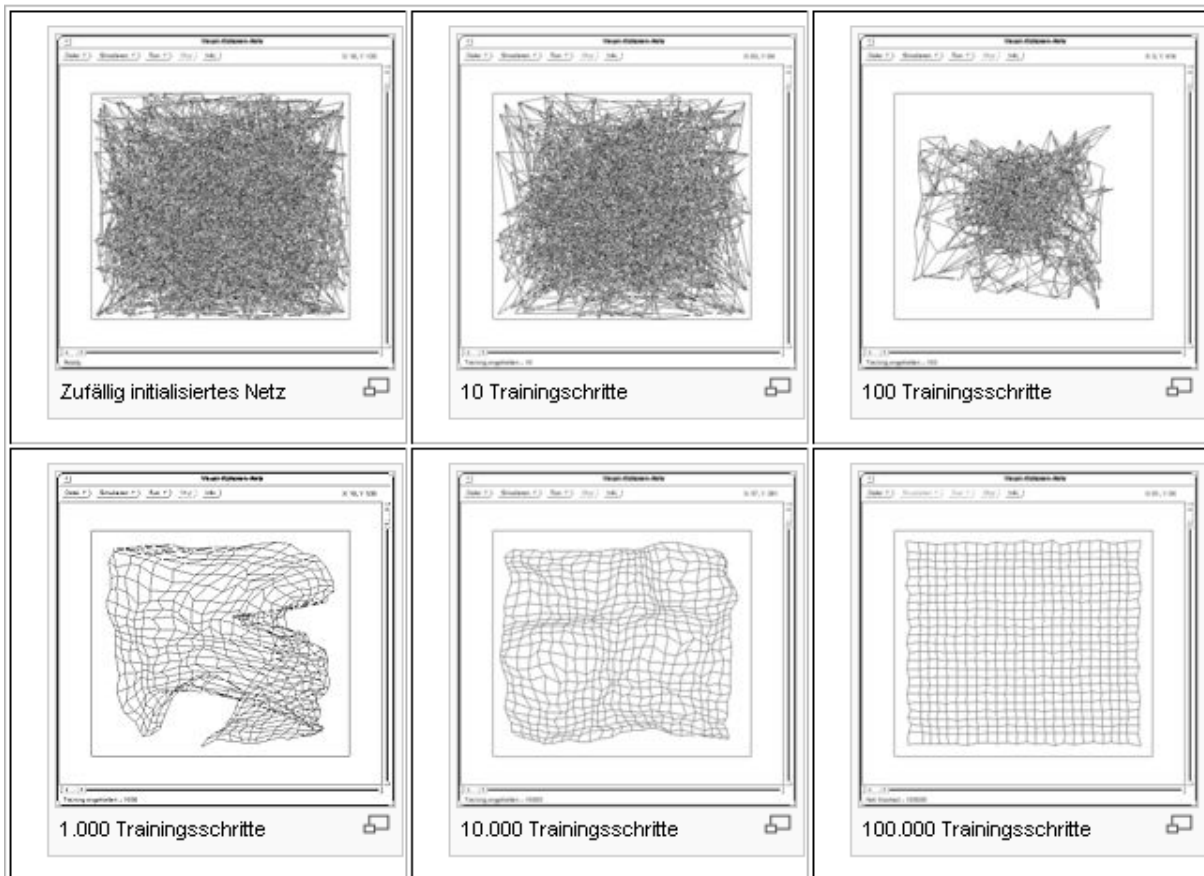


Abbildung 25: Schematische Beschreibung des Trainingsgesamtvorgangs

## 2.5 Qualitätskontrolle der Kohonenprojektion

Die Beschreibung des Trainingsvorgang in den vorangegangenen Kapiteln lässt immer noch wichtige Fragen ungeklärt: „Wie gut ist die Projektion der Eingabedaten?“ und eng damit zusammenhängend: „Wann ist das Training beendet? Wie viele *epochs* müssen für eine gute Projektion trainiert werden? Welches sind die optimalen Parameter für das Training?“ Eine allgemeingültige Lösung dieser Probleme konnte bis jetzt in der Literatur noch nicht erbracht werden.

In der ursprünglichen Publikation durch Kohonen<sup>[94]</sup> wird dieses Thema weitgehend ausgespart und durch eine visuelle Kontrolle gezeigt, dass eine 2D -> 2D Projektion erfolgreich war, siehe Abbildung 26. Als Stimuli dienen die x- und die y-Komponente von gleichverteilten Eingabewerten eines Quadrates im zweidimensionalen kartesischen Raum. Die Gewichtsvektoren der Neuronen, entsprechend der Stimuli ebenfalls eine x- und eine y-Komponente, enthalten zunächst zufällige Werte.



**Abbildung 26:** visuelle Qualitätsverbesserung einer 2D -> 2D Projektion im Verlauf des Trainings, entnommen aus <sup>[102]</sup>

Bei Auftragung dieser Werte als Punkte im xy-Graph und Verbindung benachbarter Neuronenpunkte erhält man die Grafik links oben. Im Verlauf des Trainings ergibt sich eine topologisch richtige Projektion des Quadrates der

Stimuli auf die quadratische Netzoberfläche. Anschaulich beschrieben „entfaltet sich“ die Projektion im Verlauf des Trainings, wobei die Zahl der Projektionsfehler immer weiter abnimmt. Es handelt sich hier um die einfachste denkbare Projektion, vergleichbar mit dem Training der  $n$ -Pentanhyperfläche, siehe Kapitel 2.8.2.

Alper/Bagdassarian beschreiben eine interessante Kopplung von genetischen Algorithmen und Kohonennetzen<sup>[59]</sup> zur Projektion der 3D-Koordinaten eines Proteins auf die Ebene. Dabei werden sieben empirische Parameter des Netztrainings mittels eines genetischen Algorithmus modifiziert und anhand einer Fitnessfunktion des erhaltenen Netzes optimiert. Als Fitnessfunktion dient eine spezielle Funktion, welche die topologische Richtigkeit der Projektion anhand des Erhalts von Nachbarschaftsbeziehungen zwischen Aminosäuren im 3D-Raum und Neuronen im 2D-Netzraum beschreibt. Zu jeder Aminosäure wird eine 3D-Nachbarschaft ermittelt und Punkte vergeben, wenn diese Nachbarschaft in der Kohonenprojektion erhalten bleibt. Je mehr Punkte ein fertig trainiertes Netz erhält, umso besser ist die Projektion in Bezug auf Erhalt der Topologie.

Dieses Verfahren ist in sich schlüssig, aber leider wohl nicht universell anwendbar. Alper/Bagdassarian haben nichtredundante, eindeutige Daten vorliegen. Jeder Punkt im Raum beherbergt maximal eine Aminosäure. Jede Aminosäure kommt genau einmal im Trainingssatz vor. Des Weiteren ist die Zahl der abzubildenden Zentren im Vorhinein definiert durch die Zahl der Aminosäuren, so dass die Anzahl der abzubildenden Aminosäuren im Verhältnis zu den Neuronen bei etwa 1:1 gehalten werden kann, was Skalierungseffekte in der Projektion vermeidet. Die Dauer des Trainings wird durch die iterative Wiederholung des Netztrainings im Rahmen der genetischen Algorithmen außerdem stark erhöht. Ein guter Satz Voreinstellungen ist aus praktischen Gesichtspunkten jedenfalls sehr wünschenswert.

Im allgemeinen Fall sind zur Beurteilung des Trainingserfolgs unterschiedliche Qualitätskriterien denkbar. Im Rahmen dieser Arbeit wurden folgende, unter Umständen antagonistische Ziele gesetzt:

a) Topologieerhalt:

Ähnlich zu Alper/Bagdassarian sollen Nachbarschaftsbeziehungen so weit wie möglich erhalten bleiben. Alper/Bagdassarians Fitnessfunktion ist prinzipiell wohl nur unter großem Aufwand auf den vorliegenden Fall übertragbar. Mehrfach vorkommende Strukturen (redundante Daten) führen

das Konzept der „nächsten Nachbarn im ND-Raum“ ad absurdum. Des Weiteren führt die  $O(N^2)$ -Abhängigkeit des Verfahrens in Bezug auf die Anzahl der Trainingssätze  $N$  zu einer Rechenzeitexplosion für typische sehr lange Trajektorien (Millionen Strukturen und mehr).

b) Repräsentative Abbildung:

Konformationen sollen möglichst im Raum der Trainingssätze sowie auf der Netzoberfläche den gleichen normalisierten Anteil haben. Alper/Bagdassarian haben dieses Problem nicht, da sie davon ausgehen können, dass ihre Netze im Verlauf des Trainings weitgehend eine 1:1-Zuordnung entwickeln können.

c) Vollständige Projektion:

Es sollen möglichst alle im Trainingssatz vorkommenden Strukturen akzeptabel auf die Netzoberfläche projizierbar sein. Der Widerspruch zum Kriterium b) ist offensichtlich, wenn einige Strukturen sehr oft, andere sehr selten vorkommen. Die selten vorkommenden Strukturen werden zwangsläufig irgendwann „wegtrainiert“, wenn kein freier Raum auf der Kohonenkarte vorhanden ist.

d) Maximale Trainingsgeschwindigkeit:

Das Training soll selbstverständlich möglichst schnell sein, um als Standardmethode zur Visualisierung eines Konformationsraums verwendet werden zu können.

### 2.5.1 RMSG

Diese Arbeit verfolgt als Hypothese, dass der Erhalt der Topologie a) gekoppelt ist mit einer möglichst „spannungsfreien“ Projektion. Dafür wird die mittlere quadratische Summe der inneren Gradienten der Neuronen definiert und als RMSG-Wert bezeichnet (*Root Mean Square of inner Gradients*). Je kleiner der RMSG-Wert, desto höher die Qualität der Projektion

$$(2.23) \quad \min \leftarrow RMSG = \sqrt{\frac{1}{\mu_n} \sum_{j=0}^{\mu_n} \left( d^2(N_j, N_{rechts}) + d^2(N_j, N_{rechtsunten}) + d^2(N_j, N_{unten}) \right)}$$

### 2.5.2 RMSDCF

Kriterium b) lässt sich nach Klassifizierung der Neuronen sowie der Strukturen des Trainingssatzes in  $\mu_k$  Strukturklassen ausdrücken als mittlerer quadratischer Abstand der Strukturklassenanteile an Netz und Trainingssatz, hier mit

RMSDCF-Wert abgekürzt. (*Root Mean Square Difference of Conformational Fractions*).

$$(2.24) \quad \min \leftarrow RMSDCF = \sqrt{\frac{1}{\mu_k} \sum_{k=1}^{\mu_k} \left( \frac{I_k}{\mu_a} - \frac{J_k}{\mu_n} \right)^2}$$

$I_k$  steht für die Anzahl der Trainingsstrukturen und  $J_k$  für die Anzahl der Neuronen, welche der Klasse  $k$  zugeordnet werden.

### 2.5.3 RMSDSN

Kriterium c) wird berücksichtigt, indem ein Wert der Abweichung zwischen Strukturen und repräsentativem Neuron als RMSDSN-Wert definiert wird (*Root Mean Square of Difference between Structure and Neuron*).

$$(2.25) \quad \min \leftarrow RMSDSN = \sqrt{\frac{1}{\mu_a} \sum_{i=0}^{\mu_a} d^2(A_i, N_{j_{winner}})}$$

### 2.5.4 MTS

Für eine automatische Verwendung des Verfahrens als Standardmethode ist außerdem wünschenswert, das Training mit so wenigen Schritten wie möglich durchzuführen, siehe Kriterium d).

$$(2.26) \quad MTS = epoch_{end} \cdot \mu_a$$

$epoch_{end}$  steht hier für die minimale Anzahl an Trainingsdurchläufen (*epochs*), die insgesamt trainiert werden, bis eine konstante Netzqualität, definiert durch eines der obigen Kriterien, RMSDCF, RMSG, RMSDSN erreicht ist. Sofern nicht anders angegeben, gilt im Folgenden eine konstante Qualität als erreicht, wenn die durch Abschätzung oder Extrapolation ermittelte Asymptote, z.B.  $RMSG_{end}$  bis auf 1% Abweichung erreicht wird. MTS für „*Minimum Training Steps*“ und kleine Zahlen stehen für schnelleres und damit besseres Training. Der herangezogene Qualitätsparameter wird in Klammern angegeben, z.B.  $epoch_{end}(RMSG)$  und  $MTS(RMSG)$ .

## 2.6 Grundfunktionalität von Kohonenkarten

Der Vorgang des Trainings wurde im vorhergehenden Kapitel beschrieben. Die Anwendungsmöglichkeiten eines trainierten Netzes sind bis hierhin noch relativ beschränkt. Üblicherweise wird eine einfache Zuordnungsfunktion (*lookup function*) implementiert und zur Einfärbung der Netzoberfläche mit

Eigenschaften eines Referenzsatzes an Stimuli sowie zum Auffinden und Klassifizieren beliebiger Stimuli auf der Netzoberfläche verwendet.

Neben diesen bekannten Auswertungen werden in dieser Arbeit auch neue, leistungsfähigere mehrstufige Verfahren beschrieben. Dabei kommt regelmäßig ein Satz an Grundfunktionen zum Einsatz, der in der folgenden Auflistung definiert wird.

*Lookup:*

Verfahren, das zu einer Struktur aus dem Raum der Stimuli  $A_i$  das ähnlichste Neuron  $N_j$  auffindet. Wird bereits in der Trainingsphase angewandt. Auswahl der Struktur erfolgt mit Gleichung (2.10).

*Reverse lookup:*

Rückprojektion aus dem Raum der Neuronen  $N$  in den Raum der Stimuli ergibt die Prototypen  $\mathbf{P}$  für die Neuronen, wobei jedem Neuron ein Prototyp zugeordnet wird.

$$(2.27) \quad \mathbf{P} \leftarrow \mathbf{N}$$

mit

$$(2.28) \quad \mathbf{P} = \{P_j = (\mathbf{X}^{\text{Proto}}_j) \mid j = 1, \dots, \mu_n\} \quad \mu_n : \text{Anzahl der Neuronen}$$

Am einfachsten verwendet wird als Projektionsvorschrift die Gleichung (2.29) in Analogie zu Gleichung (2.10), aber auch komplexere Verfahren wären hier denkbar.

$$(2.29) \quad \begin{aligned} i^{\text{winner}} &\leftarrow \min(d(A_i, N_j)) \quad \forall i = 1 \dots \mu_a \\ P_j &= A_{i^{\text{winner}}} \end{aligned}$$

*Kolorierung:*

Um eine aussagekräftige Netzoberfläche zu erzeugen, werden die Neuronen auf der Netzkarte eingefärbt. Einfachstes Schema ist die Verwendung von Observablen der per *reverse lookup* gefundenen Strukturen  $A_i$  nach Gleichung (2.30). In dieser Arbeit handelt es sich meist um die potentielle Energie.

$$(2.30) \quad \text{Farbwert}(N_j) = f(\text{Observable}(A_i))$$

Die Kodierung dieser Observablen im Farbschema ist prinzipiell beliebig, im Rahmen dieser Arbeit wurden verschiedene Schemata implementiert. Das hier meistverwendete Farbmodell (*landscape 3*) verläuft von blau (niedrige Werte) über grün zu gelb, dann rot und dunkelrot (hohe Werte) und ist in Gleichung (2.31) als  $(\text{Red}, \text{Green}, \text{Blue})^T$  Farbvektor definiert. Der Wert  $x$  ist dabei die zu



kodierende Observable  $Obs$ , normiert auf den Darstellungsbereich. Werte außerhalb des Darstellungsbereichs werden als Schwarz (klein) und Grau (groß) dargestellt.

(2.31)

$$\begin{array}{ccccccc} 0 < x & 0 \leq x \leq 0.2 & 0.2 < x \leq 0.4 & 0.4 < x \leq 0.6 & 0.6 < x \leq 0.8 & 0.8 < x \leq 1.0 & x > 1 \\ \begin{pmatrix} Rot \\ Grün \\ Blau \end{pmatrix} & = & \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} & \begin{pmatrix} 0 \\ 5x \\ 1 \end{pmatrix} & \begin{pmatrix} 0 \\ 1 \\ 5(0.4-x) \end{pmatrix} & \begin{pmatrix} 5(x-0.4) \\ 1 \\ 0 \end{pmatrix} & \begin{pmatrix} 1 \\ (0.8-x)/0.2 \\ 0 \end{pmatrix} & \begin{pmatrix} 1-(2.5(x-0.8)) \\ 0 \\ 0 \end{pmatrix} & \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \end{array}$$

$$\text{mit } x = \frac{Obs - Obs_{\min}}{Obs_{\max} - Obs_{\min}}$$

Die Definition aller übrigen Farbschemata ist aus der Funktion COLORREF ColorCalc(float value, minmax range) in der Datei /src/cpp/Neuro2/NeuroView.cpp der beigelegten DVD zu entnehmen. Durch manuelle Variation des Darstellungsbereichs können Details aus der Netzkarte herausgearbeitet werden.

*Gradientenanalyse:*

Die metrische Distanz der Prototypstrukturen zweier benachbarter Neuronen  $N_j$  und  $N_k$  wird in Analogie zu Gleichung (2.9) definiert durch Gleichung (2.32).

$$(2.32) \quad d(P_j, P_k)^2 = \sum_{p=1}^M d_{euclid}(X_{j,p}^{Proto}, X_{k,p}^{Proto})^2$$

Durch Variation der Dicke der Trennlinie zwischen diesen Neuronen auf der Netzkarte unter Berücksichtigung der metrischen Distanz der Prototypstrukturen ergibt sich ein visuelles *Clustering*. Ebenso lassen sich die metrischen Distanzen der Neuronen direkt nach Gleichung (2.8) heranziehen. Im Rahmen dieser Arbeit wird das einfache zweistufige Schema aus Gleichung (2.33) verwendet.

$$Liniendicke(N_j, N_k) = \begin{cases} 0 & \forall \quad 0 \leq g < 0.5 \\ 1 & \forall \quad 0.5 \leq g \leq 1 \end{cases}$$

mit :

$$(2.33) \quad g = \frac{d(X_j^{Proto}, X_k^{Proto}) - grad_{\min}}{grad_{\max} - grad_{\min}},$$

$$grad_{\min} = \text{Min}(d(X_j^{Proto}, X_k^{Proto})) \quad \forall N_j \text{ Nachbar von } N_k$$

$$grad_{\max} = \text{Max}(d(X_j^{Proto}, X_k^{Proto})) \quad \forall N_j \text{ Nachbar von } N_k$$

*Trajektorienanalyse:*

Die Trajektorienanalyse besteht aus einem *Lookup* für jedes  $A_i$  einer Sequenz von Stimuli (Konformationen). Die erhaltene Sequenz an zugehörigen Neuronen wird auf der Netzoberfläche verbunden. Diese Analysemethode offenbart, welcher Bereich der Hyperfläche des trainierten Netzes von der aktuellen Sequenz (Trajektorie) an Stimuli untersucht wird. Eine weitere Anwendung liegt im Verfolgen von Moleküldynamiksimulationen und einfachen Auffinden von Konformationswechseln.

*Mehrstufenoptimierung (Multipass):*

Iteratives Optimieren der per *reverse lookup* gefundenen Prototypstrukturen der Neuronen mit unterschiedlichen Rechenverfahren. Dabei werden die bereits optimierten Prototypen als Startstrukturen für eine erneute Optimierung verwendet.

$$(2.34) \quad \begin{array}{l} \mathbf{Opt} \leftarrow \mathbf{P} \\ \mathbf{P}' \equiv \mathbf{Opt} \\ \mathbf{Opt}' \leftarrow \mathbf{P}' \\ \vdots \end{array}$$

Kolorierung und Gradientenanalyse lassen sich auf der Netzkarte für jede Iteration durchführen. Eine typische Anwendung zur Projektion von Konformerensembles besteht aus den folgenden Schritten:

1. Durchführen eines Moleküldynamik-Laufs (oder einer anderen Methode der Konformerengenerierung) und Speichern der erhaltenen Trajektorie.
2. Auswahl der Geometrieparameter für das Kohonennetz. Ermitteln des Wertebereichs zur Normalisierung.
3. Training des Kohonennetz mit obigen Parametern. Die Gewichtsvektoren  $\mathbf{W}$  der  $\mu_n$  Neuronen können als Prototypen im Raum der trainierten Strukturen angesehen werden.
4. Zuordnung von Prototypstrukturen zu den Neuronen aus dem Raum der Stimuli (*reverse lookup*).
5. Eine partielle Geometrieoptimierung der erhaltenen Prototypstrukturen unter Anwendung von *restraints*. Als *restraints* verwendet werden die Gewichte der Neuronen, zurückgerechnet in den Raum der Geometrieparameter. Alle anderen Freiheitsgrade werden frei optimiert.

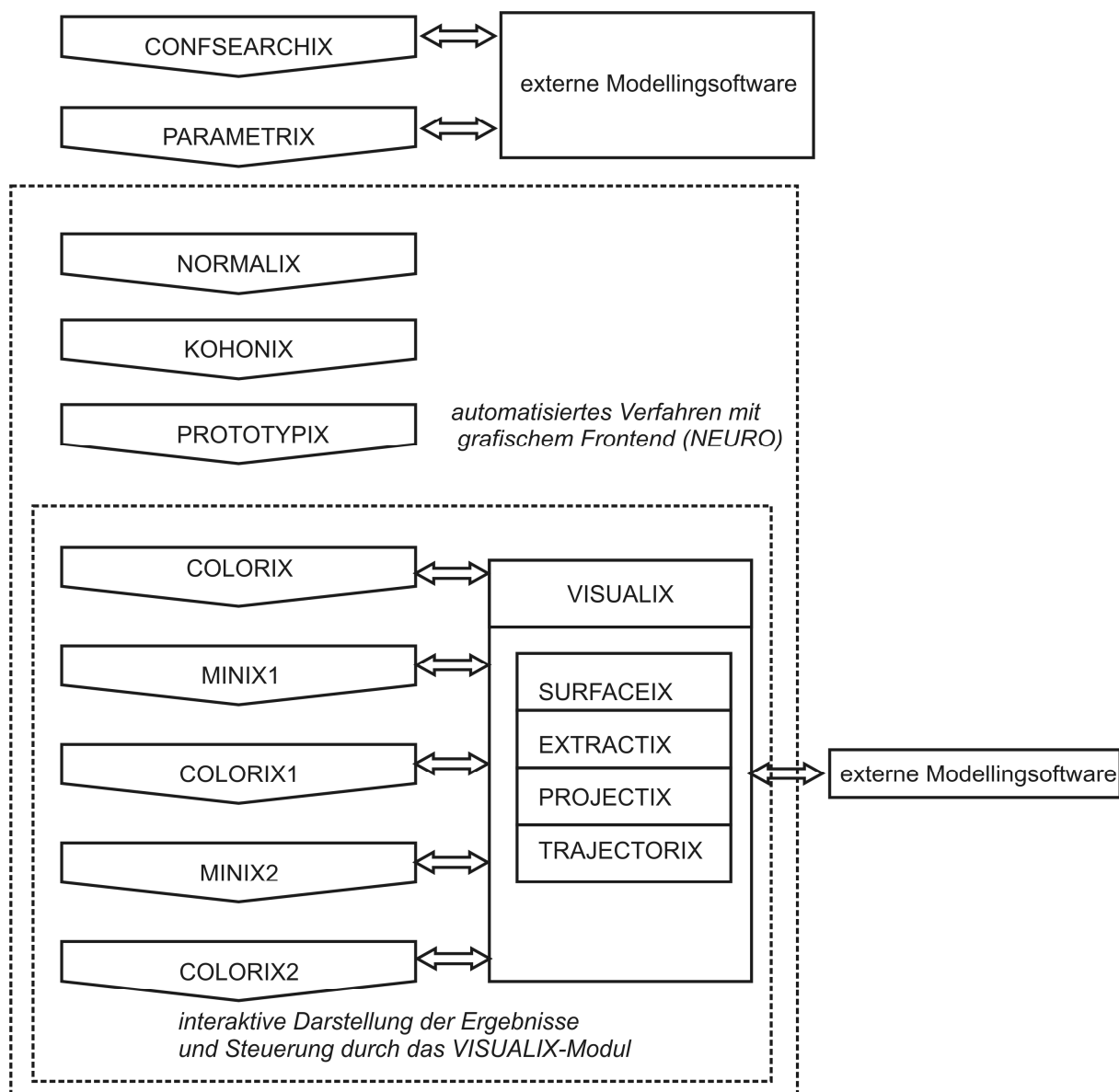
6. Kolorierung der Neuronen unter Verwendung der potentiellen Energie der partiell optimierten Strukturen aus dem vorherigen Schritt.
7. Mehrstufenoptimierung (*2nd pass*): freie Optimierung ohne *restraints* der in Schritt 5 voroptimierten Strukturen.
8. Visuelles *Clustering* durch Gradientenanalyse der Netzoberfläche (2.8). Verwendung der optimierten Strukturen **Opt**(2.34), aus Schritt 5 zur Kolorierung und den erneut optimierten Strukturen aus Schritt 7 zur Gradientenanalyse.
9. Eintragen der Trajektorie des ursprünglichen Konformerensemble (nur für sequentielle Konformationsmethoden wie MD und MC) auf der Netzoberfläche.

## 2.7 Die NEURO-Programmsuite

Die NEURO-Programmsuite wurde im Rahmen dieser Arbeit entwickelt, um die zuvor besprochenen Analysemethoden zu implementieren. Sie besteht aus einem ganzen Satz größtenteils unabhängiger Programme und Skripte. Für einen zusammenfassenden Überblick ist eine Darstellung des Fließschemas der Vorgehensweise jedoch zunächst besser geeignet als eine Beschreibung der Programme. Die allgemeine Beschreibung des Verfahrens geschieht deshalb zunächst im Kapitel 2.7.1 anhand des Fließschemas in Abbildung 27. Einzelheiten der Implementierung und der Benutzung werden im anschließenden Kapitel 2.7.2 sowie in den aufgeführten Beispielanwendungen diskutiert.

### 2.7.1 Module als Organisationsprinzip

Die Anwendung der NEURO-Programmsuite wird nach Abbildung 27 in logische Module unterteilt, die zunächst unabhängig von der aktuellen Implementierung sind. Ein Modul wird in diesem Zusammenhang definiert als *Blackbox*, die ein bestimmtes Problem bzw. eine bestimmte Aufgabe löst. Ihre Beschreibung geschieht in den nachfolgenden Unterkapiteln.



**Abbildung 27** Fließschema der Analyse von Konformerensembles mit der Programmsuite NEURO

### 2.7.1.1 CONFSEARCHIX

Die Aufgabe des Moduls CONFSEARCHIX ist die Erstellung des zu untersuchenden Konformerensembles nach einer beliebigen der in 1.2 angeführten Methoden. In dieser Arbeit werden die Programme HYPERCHEM und SYBYL samt Ihrer diversen MD-, MC- und *conformational search* Routinen verwendet.

### 2.7.1.2 PARAMETRIX

Das Training des Netzes geschieht in einem Parameterraum, der durch den Anwender definiert wird. Damit wird bereits eine Vorreduktion des Parameterraums durchgeführt. Grundsätzliches zur Auswahl dieser Parameter wird in Kapitel 2.4 beschrieben. Die Kombination von den Anwender interessierenden Struktureigenschaften und strukturbestimmenden internen Koordinaten ergibt auf den jeweiligen Anwendungszweck „maßgeschneiderte“ Karten.

### **2.7.1.3 NORMALIX**

Ermittelt werden die minimalen und maximalen Werte der obigen Parameter, wie sie im Ensemble vorkommen. Diese werden benötigt zur Normalisierung der Eingabedaten. Ab hier werden die logischen Module durch die NEURO Programmsuite abgedeckt.

### **2.7.1.4 KOHONIX**

Dieses Modul enthält die Extraktion der zu trainierenden Parameter aus dem Konformerensemble sowie Erzeugung und Training eines Kohonennetzes mit den normalisierten Werten.

### **2.7.1.5 PROTOTYPIX**

Ordnet den einzelnen Neuronen eine Prototypstruktur aus dem ursprünglichen Konformerensemble mittels *reverse lookup* zu.

### **2.7.1.6 COLORIX**

Extrahiert die potentielle Energie der Prototypen als Observable zur Einfärbung der Netzoberfläche im VISUALIX-Modul. Bereitet die Gradientenanalyse vor.

### **2.7.1.7 MINIX1**

Erster Schritt einer Mehrstufenoptimierung. Führt eine Geometrieoptimierung der Prototypstruktur jedes Neurons, erhalten aus dem COLORIX-Modul, durch. Eine kritische Verbesserung zu literaturbekannten Verfahren besteht in der Rückprojektion der Neurongewichte als zusätzliche *restraints* für die Optimierung. Die Gewichte des Neurons repräsentieren das Ensemble der ursprünglichen Stimuli, welche durch das Neuron abgedeckt werden, weitaus besser, als es eine einzelne Prototypstruktur vermag. Die MINIX-Module verwenden im Rahmen der NEURO-SUITE das Programm HYPERCHEM als MM-backend.

### **2.7.1.8 COLORIX1**

Extrahiert die potentielle Energie der im Modul MINIX1 unter Anwendung von *restraints* optimierten Prototypstrukturen zur Einfärbung der Netzoberfläche im VISUALIX-Modul. Bereitet die Gradientenanalyse vor.

### **2.7.1.9 MINIX2**

Zweiter Teil einer Mehrstufenoptimierung. Führt eine *restraint*-freie Geometrieoptimierung der in Modul MINIX1 voroptimierten Prototypstrukturen durch.

### **2.7.1.10 COLORIX2**

Extrahiert die potentielle Energie der im Modul MINIX2 optimierten Prototypstrukturen als Observable zur Einfärbung der Netzoberfläche im VISUALIX-Modul. Bereitet die Gradientenanalyse vor.

### **2.7.1.11 VISUALIX**

Die Visualisierung und interaktive Bearbeitung der erhaltenen Ergebnisse sowie das Starten und Überwachen der Rechnungen erfolgt über das Modul VISUALIX. Dabei handelt es sich um eine Sammlung unterschiedlicher Methoden zur Darstellung der Ergebnisse der oben angeführten eher rechnerisch orientierten Module. Diese werden im Folgenden aufgelistet.

### **2.7.1.12 SURFACEIX**

Stellt die Kohonenkarte als 2-dimensionales Raster dar. Die einzelnen Neuronen werden durch die Werte der Observablen aus den selektierten COLORIX-Modulen eingefärbt. Des Weiteren können die Interneurongradienten, welche die metrische Distanz der Prototypstrukturen symbolisieren, kodiert in Form unterschiedlich dicker Trennlinien zwischen den Neuronen dargestellt werden.

### **2.7.1.13 EXTRACTIX**

Anzeige der Prototypstrukturen der Neuronen, „roh“ oder nach Optimierung in einem der MINIX Module. Im Rahmen dieser Arbeit wird als externer Molekülbetrachter das Programm HYPERCHEM verwendet, welches per DDE (*dynamic data exchange*) fast vollständig ferngesteuert werden kann.

### **2.7.1.14 PROJECTIX**

Zu einer von außen (HYPERCHEM) vorgegebenen Struktur wird das „repräsentativste“ Neuron gefunden und mit einem Punkt auf der Kohonenkarte markiert. Implementiert also eine *lookup*-Operation.

### **2.7.1.15 TRAJECTORIX**

Führt das Modul PROJECTIX für jede Struktur eines Satzes an Stimuli und verbindet die Punkte auf der Netzoberfläche.

### 2.7.2 Übersicht der Kernprogramme:

Im Rahmen der NEURO-SUITE werden die in Kapitel 2.7.1 beschriebenen Module durch einzelne Programme implementiert. Als Übersicht, welche Programme in welchem Teil der Analyse zu verwenden sind, dient Tabelle 9.

**Tabelle 9:** Zuordnung der Programme der NEURO-SUITE zu den Modulen. In eckigen Klammern ist das Programm angegeben, welches üblicherweise den Aufruf des Programms vornimmt.

Modul	Programme
CONFSEARCHIX	HYPERCHEM, SYBYL, PCMODEL
PARAMETRIX	HYPERCHEM LINTORS.TK, RINGTORS.TK, MAKE_DISTANCE_MATRIX.TK
NORMALIX	ANALYSE.EXE – [NEURO.EXE]
KOHONIX	NN.EXE – [NEURO.EXE]
PROTOTYPIX	ASSIGN.EXE – [NEURO.EXE]
COLORIX, ~1, ~2	MINIMIZE.EXE / ASSIGN.EXE / MAKE2ND.EXE – [NEURO.EXE]
MINIX1, ~2	MINIMIZE.EXE – [NEURO.EXE]
VISUALIX	NEURO.EXE
SURFACEIX	
EXTRACTIX	
PROJECTIX	
TRAJECTORIX	

Im Rahmen der Verwendung der NEURO-SUITE fallen dabei eine Vielzahl von Dateien an. Einen kurzen Überblick über den Inhalt der erzeugten Dateien vermittelt Tabelle 10.

**Tabelle 10:** Dateien der NEURO-SUITE

Dateiname	Inhalt
beispiel.hin	HYPERCHEM Datei mit dem molekularen System. Enthält die zu trainierenden Parameter in Form von <i>selections</i> .
beispiel.snp	Strukturen der zu trainierenden Konformationen.
beispiel.dis	Trainingsparameter mit Wertebereich über beispiel.snp.
beispiel.net	Gewichtsvektoren der einzelnen Neuronen, samt Zuordnung der ähnlichsten Strukturen aus beispiel.snp.
beispiel.net.log	Statistik über den Trainingsverlauf des Netzes.
beispiel.sav	wie beispiel.net, im Binärformat. Obsolet.
beispiel.scr	Indizes in beispiel.snp der zu optimierende Strukturen für MINIX1.

beispiel.slg	Optimierte Strukturen aus MINIX1.
beispiel.slg.log	Statistik über den Verlauf von MINIX1.
beispiel.eng	den Neuronen zugeordnete Energien (minimiert und nicht minimiert).
beispiel.2nd.net	wie beispiel.net, mit Zuordnungen der Neuronen passend für MINIX2
beispiel.2nd.scr	Indizes in beispiel.slg der zu optimierenden Strukturen für MINIX2.
beispiel.2nd.slg	Optimierte Strukturen aus MINIX2.
beispiel.2nd.slg.log	Statistik über den Verlauf von MINIX2.
beispiel.2nd.eng	den Neuronen zugeordnete Energien (minimiert und 2nd pass minimiert).

Die Abhängigkeiten der Programme voneinander lassen sich am Datenflussdiagramm Abbildung 28 erkennen.



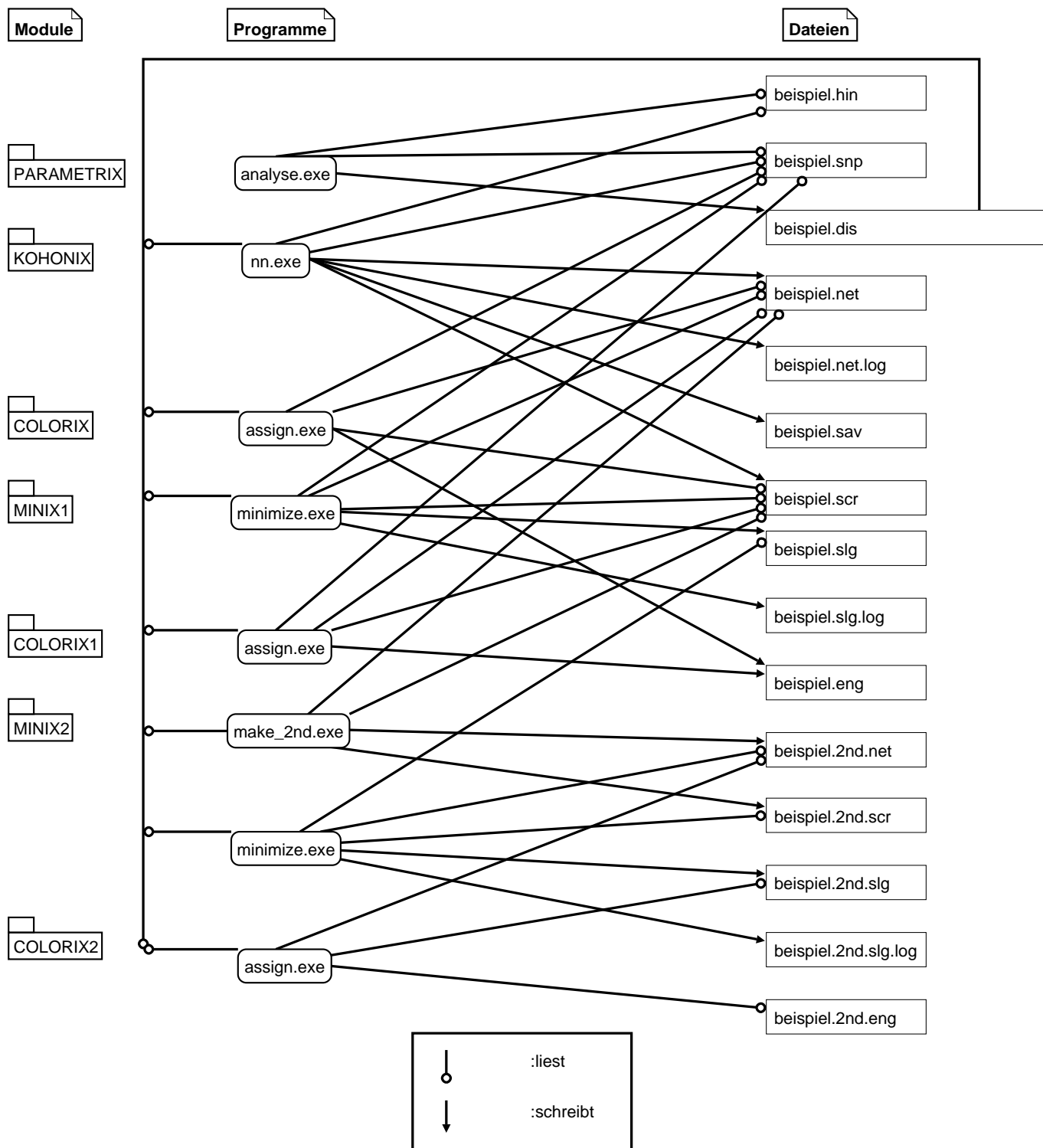


Abbildung 28: Datenflussdiagramm der NEURO Programmsuite

In folgenden Kapiteln wird die Verwendung der Programme ausführlich und mit *Screenshots* beschrieben.

## 2.8 Evaluierung des Netztrainings

Zur Verifizierung der neuronalen Netze werden zunächst kleine Moleküle herangezogen. Deren wohlbekannte Hyperflächen erlauben es, die Parameter des Trainingsvorgangs kontrolliert zu optimieren.

### 2.8.1 *n*-Butan

Die Hyperfläche des Butans ist relativ einfach und schon recht lange bekannt. Es gibt eine *anti*- und zwei *gauche*-Konformationen, *gauche*<sup>+</sup> und *gauche*<sup>-</sup>, die sich zueinander wie Bild und Spiegelbild verhalten, also Enantiomere sind.

**Tabelle 11: Nomenklatur der Diederwinkel**

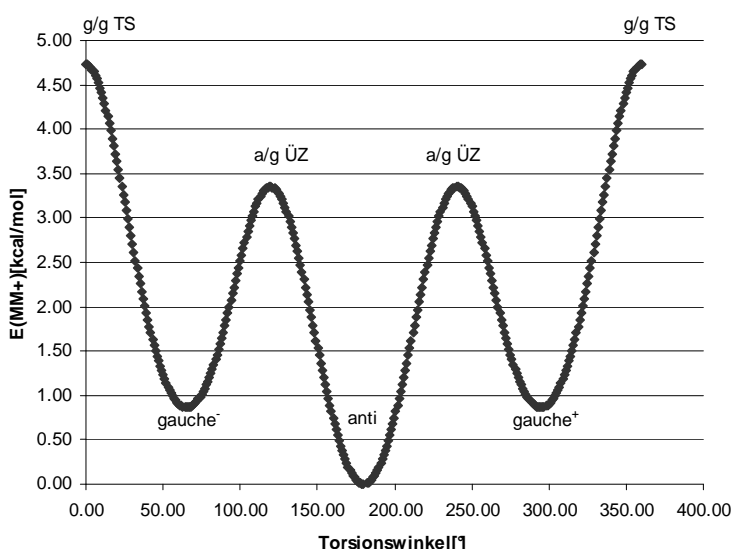
Name	Wertebereich
<i>anti</i> (a)	120°..240
<i>gauche</i> <sup>+</sup> (g <sup>+</sup> )	0°..120°
<i>gauche</i> <sup>-</sup> (g <sup>-</sup> )	240°..360°

Experimentelle Daten ergeben nach Eliel,<sup>[26]</sup> siehe Abbildung 4, S. 11, eine Energiedifferenz zwischen *anti*- und *gauche*-Minimum in Höhe von 0.9 kcal/mol und eine Übergangsbarriere in Höhe von 3.6 kcal/mol zwischen *gauche*- und *anti*-Minimum respektive 4.5..6.1 kcal/mol zwischen den *gauche*-Isomeren. Auf MM<sup>+</sup> - Kraftfeldniveau können diese Werte ausgezeichnet nachvollzogen werden, siehe Tabelle 12.

**Tabelle 12: Energiehyperfläche des Butans, Kraftfeldniveau (MM<sup>+</sup>)<sup>[103]</sup>**

Energie[kcal/mol]	<i>anti</i>	<i>gauche</i>	a/g-ÜZ	a/g-ÜZ
E <sub>exp</sub> nach Eliel[31]	0	0.9	3.6	4.5 .. 6.1
E <sub>MM<sup>+</sup></sub> [121]	0	0.86	3.35	3.83

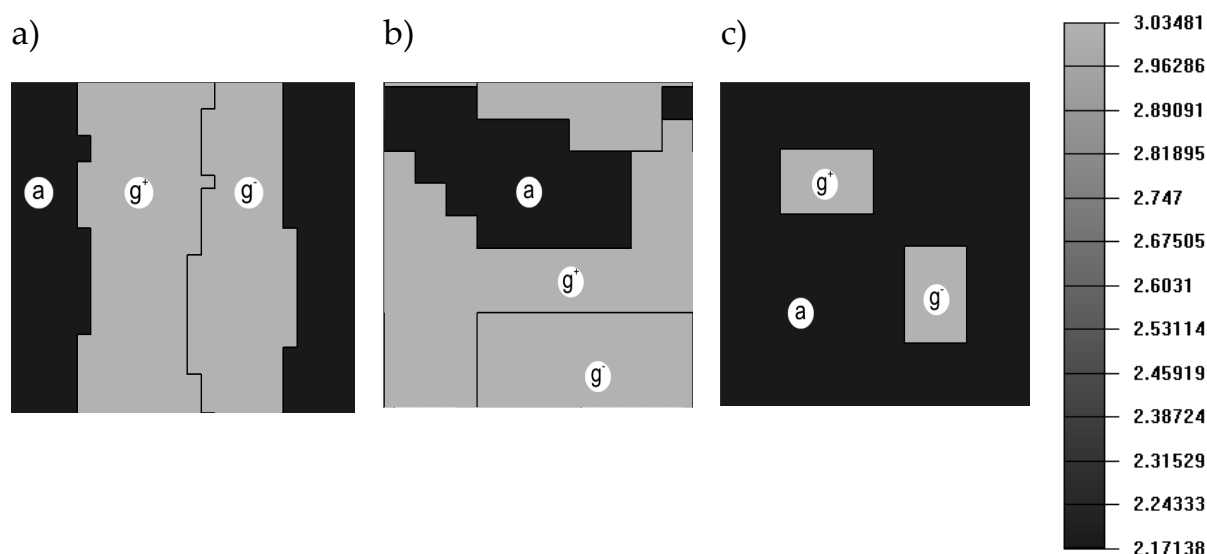
Auch der Plot der Energien zeigt die Übereinstimmung dieser Methode mit den experimentellen Daten, vergleiche Abbildung 4, S. 11 mit Abbildung 29.



**Abbildung 29: grafische Auftragung der Energiehyperfläche des Butans auf Kraftfeldniveau(MM<sup>+</sup>)**

Die Energiehyperfläche ist insofern interessant, als dass sie prinzipiell mit einem Freiheitsgrad beschrieben werden kann, während das trainierte Koho-

ennetz zweidimensional ist und man es mit einer Projektion aus einem niederdimensionalen Raum in einen höherdimensionalen Raum zu tun hat. Gefunden wurden vom Netz je nach Bedingungen folgende Prototypen einer topologisch korrekten vollständigen Hyperfläche, siehe Abbildung 30.



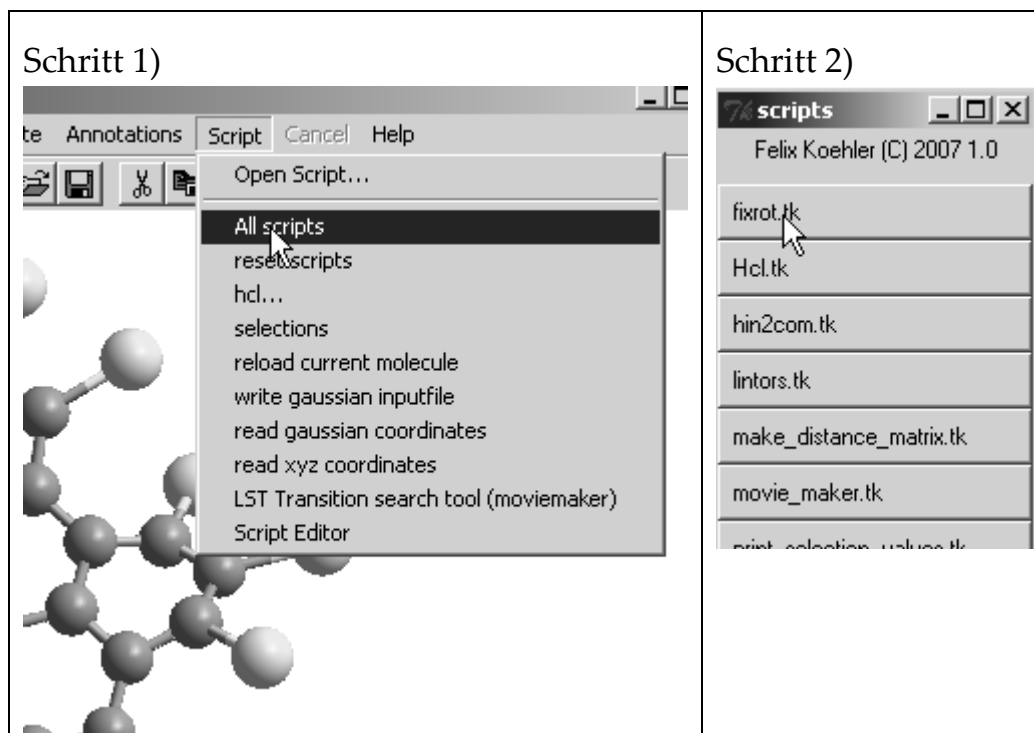
**Abbildung 30:** Projektionen der Hyperfläche des Butans mit Energieskala (in kcal/mol)

### 2.8.1.1 Tutorial: MD-Rechnungen mit HYPERCHEM

Um an Daten zum Training des neuronalen Netzes zu gelangen, muss zunächst ein Satz Konformerer in einer *snapshotfile* Datei (snp-Datei) des Programms HYPERCHEM erzeugt werden. Eine einfache und effiziente Möglichkeit besteht in der Durchführung einer Moleküldynamikrechnung mit dem Programm HYPERCHEM. Im Rahmen dieser Doktorarbeit hat sich gezeigt, dass MD-Simulationen im Programm HYPERCHEM mit einem schweren *Bug* behaftet sind. Die Moleküle geraten in Rotation, wodurch die innere Energie des Moleküls zum großen Teil gebunden wird. Außerdem wird das molekulare System aufgrund der Fliehkräfte in eine gestreckte Konformation überführt, was den erreichbaren Konformationsraum einschränkt. Da es sich bei Kraftfeldrechnungen um klassisch mechanische Simulationen handelt, kann die Quantelung der Rotationsenergien keinesfalls korrekt berücksichtigt werden. Eine Methode, durch Korrektur der Geschwindigkeitsvektoren der Atome während der Simulation trotzdem verwertbare Trajektorien zu erhalten besteht in der Verwendung des erstellten Hilfsskripts FIXROT.TK (siehe Kapitel 5.2.1, S. 335).

Das Skript wird im Rahmen der modifizierten HYPERCHEM - Umgebung (*chem.scr*) durch Aufruf des Menüpunkts „Script/All scripts ...“ zugänglich gemacht. Dies ist eine allgemeine Vorgehensweise für Erweiterungen des

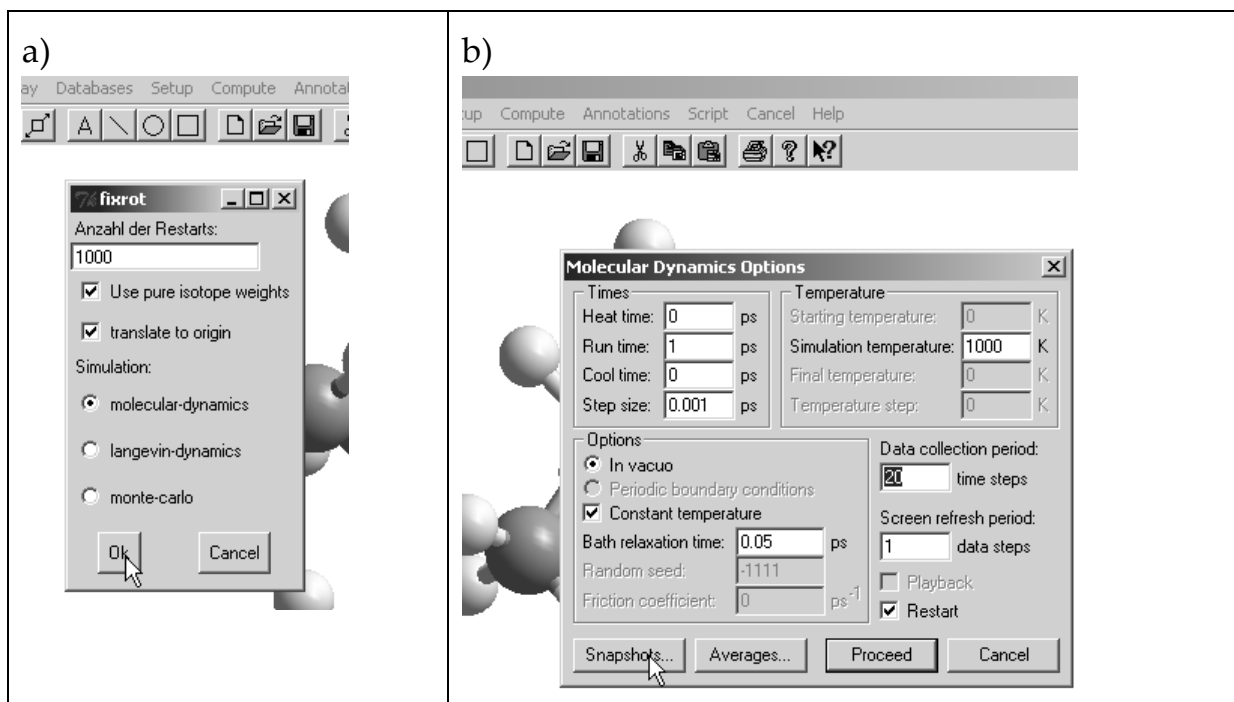
Programms HYPERCHEM, welche im Rahmen dieser Doktorarbeit erstellt wurden und wird an dieser Stelle einmal ausführlich in Abbildung 31 beschrieben.



**Abbildung 31: Tutorial - Einbindung eine Hilfsskripts in das Programm HYPERCHEM**

Der Fixrot Menüeintrag ermöglicht jetzt den Aufruf des Hilfsskripts FIXROT.TK. Dieses ruft HYPERCHEMS eingebauten „Compute/Molecular Dynamics“ Dialog (optional auch „Compute/Langevin Dynamics“ oder „Compute/Monte-Carlo“) mehrfach hintereinander auf, korrigiert die Geschwindigkeitsvektoren mittels des Programms FIX\_ROTATION.EXE (Kapitel 5.3.3) und verbindet die erzeugten snp-Dateien mittels des Programms JSNAPS.EXE (siehe Kapitel 5.3.9). Die *Restart* - Option (Abbildung 32 b) sollte dabei aktiviert sein.

Die Gesamtsimulationslänge berechnet sich aus der Anzahl der *Restarts* in Abbildung 32a) multipliziert mit der *Run Time* des eingebauten Dialogs in Abbildung 32b). Der *Snapshots...* Dialog muss dabei zwingend verwendet werden, um die Simulation zu speichern und nicht nur im Speicher durchzuführen. Da *Heat Time* und *Cool Time* zwingend Null sein müssen, sollte mit dem System unmittelbar zuvor eine entsprechende Simulation ohne Speicherung zwecks Equilibrierung durchgeführt worden sein.



**Abbildung 32: Tutorial - Aufruf des FIXROT.TK Skriptes**

Aus obigem Beispiel (Kraftfeld MM+) erhält man eine snp-Datei, hier mit dem Namen `butan_tutorial.snp` mit einer Gesamtsimulationslänge von 1 ns (1000x1 ps). Bei einer *data collection period* von 10 s<sup>-1</sup> entspricht dies 100000 hochredundanten Eingabestrukturen, welche ca. 51MB Speicherplatz belegen.

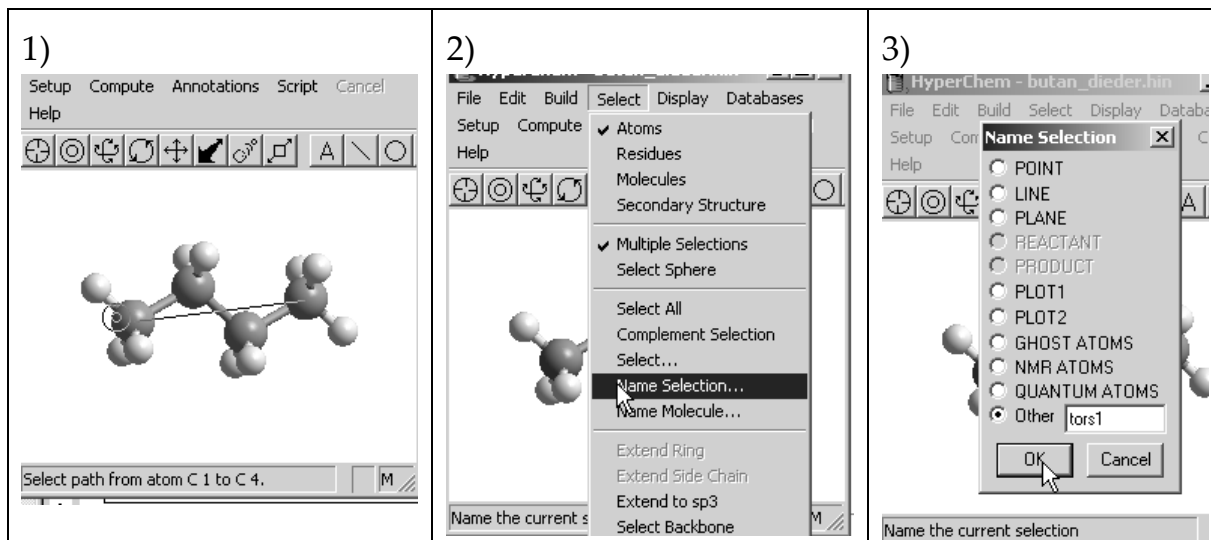
### 2.8.1.2 Tutorial: Training einer Kohonenkarte

Mit den obigen Moleküldynamik-Daten wird im Folgenden exemplarisch die Verwendung der NEURO-SUITE zum Training einer Kohonenkarte vorgestellt. Dieses Vorgehen wird in den weiteren Beispielen als *Modus Operandi*, soweit nicht anders angegeben, vorausgesetzt. Das Vorgehen orientiert sich an Abbildung 27, S. 66.

#### PARAMETRIX (Definition der Trainingsparameter)

Die im Netz trainierbaren Parameter sind interne Koordinaten des molekularen Systems: Abstände, Winkel, Diederwinkel. Diese können durch HYPERCHEM *selections* dargestellt werden, was es ermöglicht, mit HYPERCHEM als grafischem Editor die Auswahl der Trainingsparameter durchzuführen.

Dieser Vorgang ist in Abbildung 33 dargestellt.



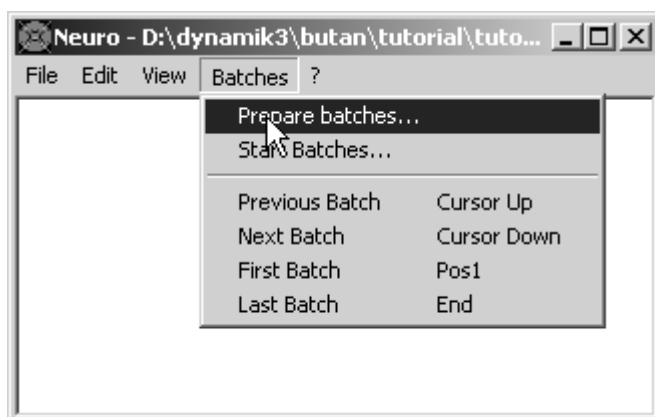
**Abbildung 33: Tutorial - Definition von selections in HYPERCHEM**

Diese Datei mit der *selection* des Diederwinkels, aufgespannt durch die Atome (1, 2, 3, 4), wird unter dem Dateinamen butan\_dieder.hin gespeichert als Definition eines Trainingsatzes, der genau aus diesem einen Diederwinkel besteht.

Für die Erstellung von Sätzen an Trainingsparametern gibt es außerdem die Hilfsskripte LINTORS.TK (Kapitel 5.2.4), RINGTORS.TK (Kapitel 5.2.8) sowie MAKE\_DISTANCE\_MATRIX.TK (Kapitel 5.2.5). Diese werden eingebunden gemäß Abbildung 31. Ein Trainingsatz der Distanzmatrix der Schweratome wird unter dem Dateinamen butan\_distances.hin gespeichert. Der Trainingsatz ist integraler Bestandteil der Auswertung und muss zu jeder Netzkarte angegeben werden.

NORMALIX:

Das Programm NEURO.EXE (siehe Kapitel 5.3.14) dient als grafisches *front end* für die nun folgenden Schritte. Gestartet wird mit einem leeren Projekt, das Planen der durchzuführenden Auswertungen geschieht im Menüpunkt „Batches/Prepare Batches ...“ (siehe Abbildung 34)



**Abbildung 34: Tutorial - Planen eines NEURO - Projekts**

Für das Modul NORMALIX werden eine Beschreibung der Auswertung (Abbildung 35a), die HYPERCHEM-Datei mit den Trainingsparametern (Abbildung 35b) als *selections*, die analog ausgewählte snp-Datei sowie der Name der zu erzeugenden Parameterdatei (Abbildung 35c) benötigt.

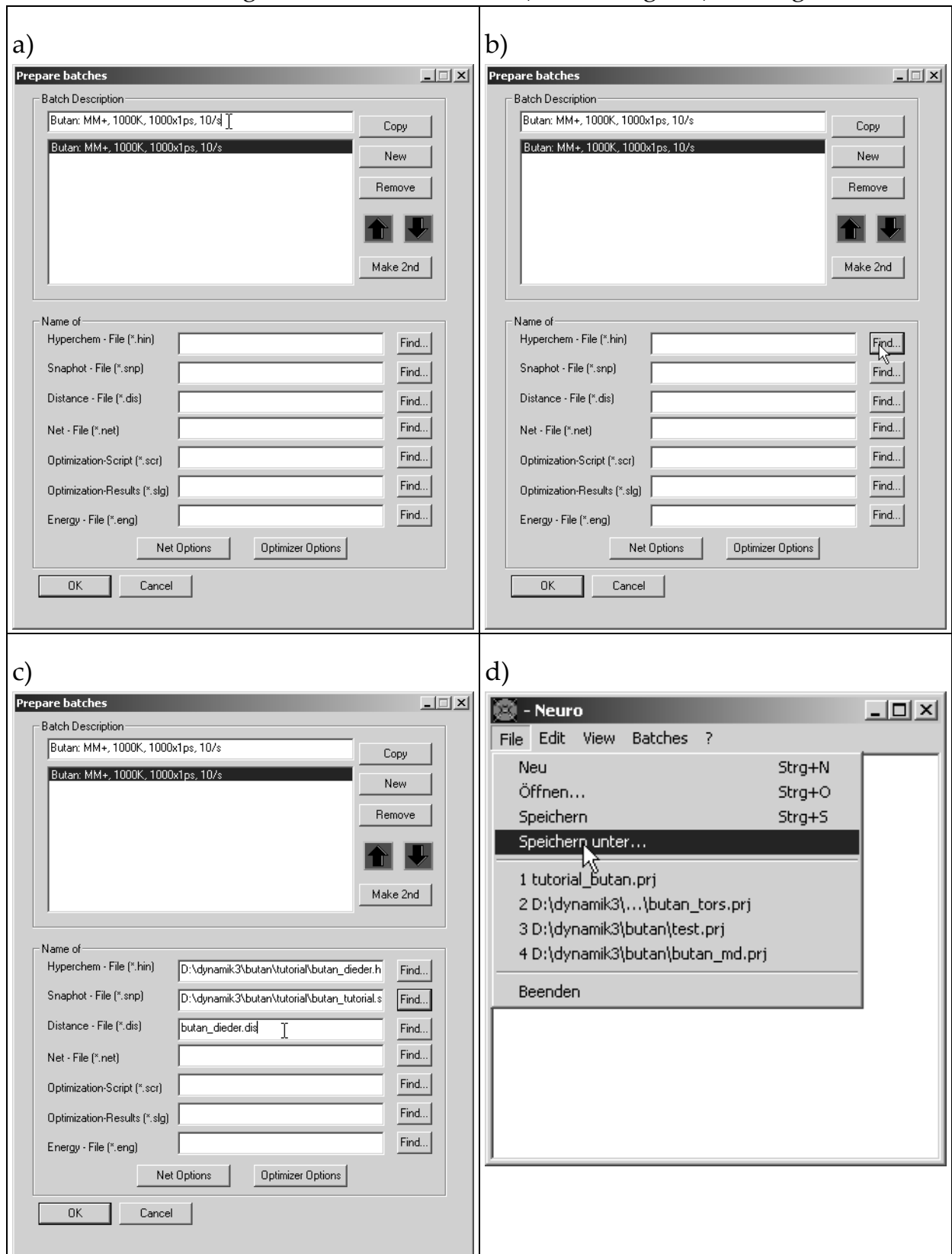


Abbildung 35: Tutorial NORMALIX, Einrichtung der Rechnungen

Die Einstellungen zu den Rechnungen (*Batches*) werden über den Menüpunkt „File/Save As...“ bzw. „File/Speichern unter...“ als Klartext-Projektdatei gespeichert und sind damit später wieder abrufbar. Es kann jetzt das Modul NORMALIX gestartet werden. Dazu wird der Menüpunkt „Batches/Start Batches“ aufgerufen (siehe Abbildung 36a). Das Modul NORMALIX wird selektiert und auf den aktuellen Datensatz (*Batch*) angewandt (siehe Abbildung 36b). Der in Abbildung 36c dargestellte Dialog zeigt den aufgerufenen Prozess *Analyse.exe* samt Kommandozeile mit Möglichkeit zum Abbruch an. Nach Abschluss der Rechnung steht eine *distance* Datei (*butan\_dieder.dis*) mit extrahierten Parametern samt Wertebereich zur Normalisierung bereit.

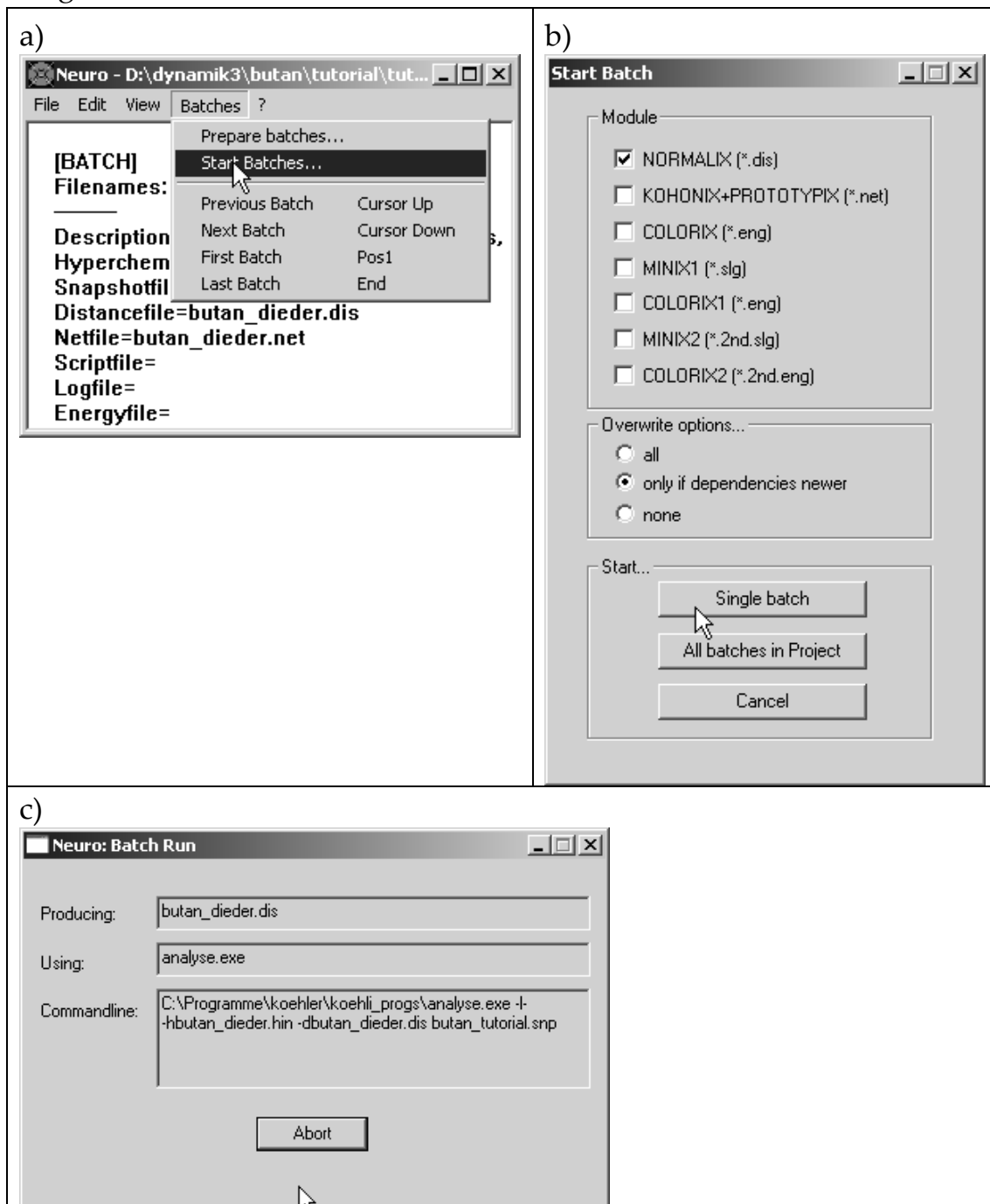
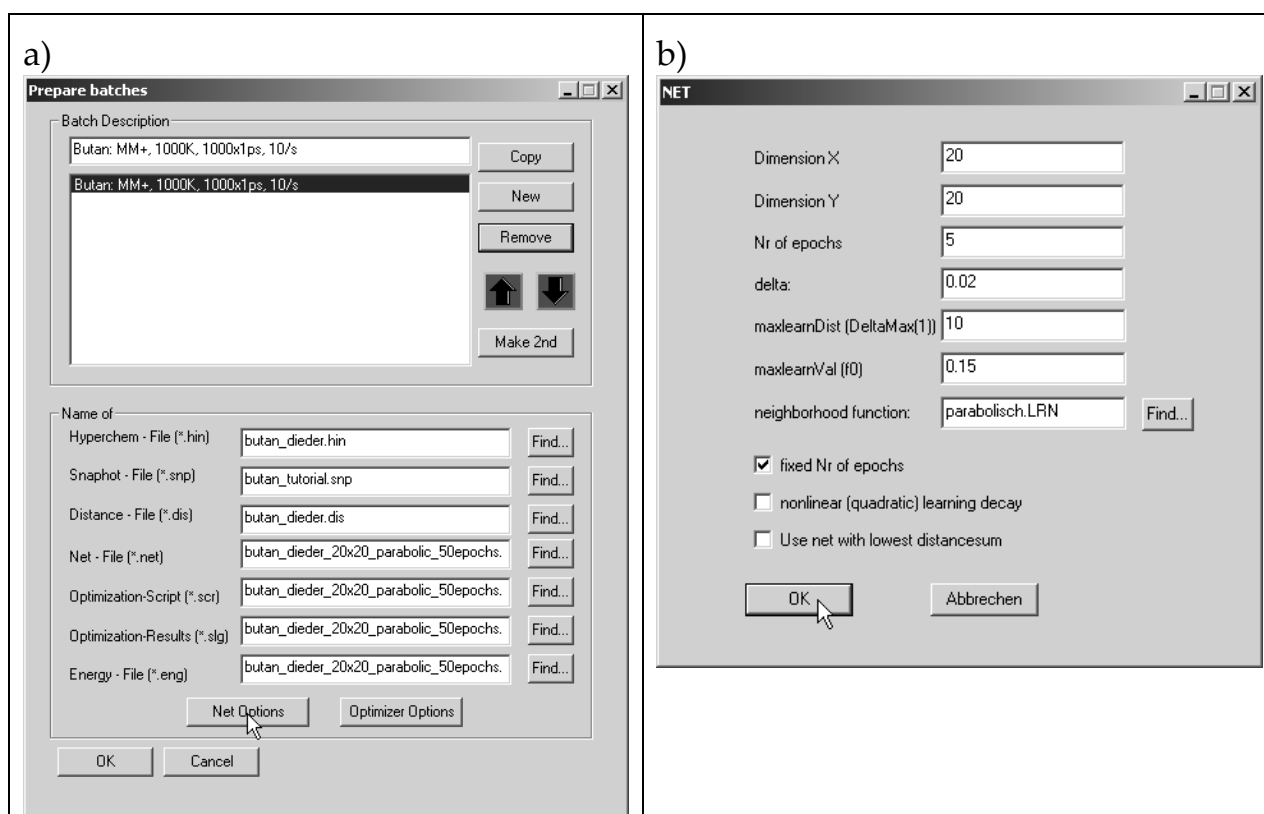


Abbildung 36: Tutorial NORMALIX, Start des Moduls



## KOHONIX/PROTOTYPIX:

Diese Module sind aus Effizienzgründen zu einem Durchgang vereint. Es wird wiederum nach Abbildung 34 der Menüpunkt „Batches/Prepare Batches“ aufgerufen um die benötigten Informationen zu den Netzparametern beziehungsweise den Dateinamen anzugeben. Als Name der Netzdatei samt Indizes der zugeordneten Prototypen wird in diesem Tutorial der Dateiname `butan_dieder_20x20_parabolic_50epochs.net` verwandt. Alle weiteren Verarbeitungsschritte sind direkt von dieser Datei abhängig und erhalten hier den gleichen Namen, modulo Endung. Damit sind von den Dateinamen her alle Arbeitsschritte vorbereitet.



**Abbildung 37:** Tutorial KOHONIX, Wahl der Netzparameter

Die Angabe der Netzparameter (*Net Options*) geschieht wie in Abbildung 37 illustriert. Diese Netzparameter sind integraler Bestandteil der Auswertung und müssen zu jeder Netzkarte angegeben werden. In folgender Tabelle 13 wird eine Erläuterung der Parameter des NET-Dialogs vorgenommen.

**Tabelle 13:** Tutorial KOHONIX, tabellarische Auflistung der verwendeten Netzparameter mit kurzer Erläuterung.

Parameter	Wert	Erläuterung
DimensionX	20	Anzahl der Neuronen in X-Richtung
DimensionY	20	Anzahl der Neuronen in Y-Richtung

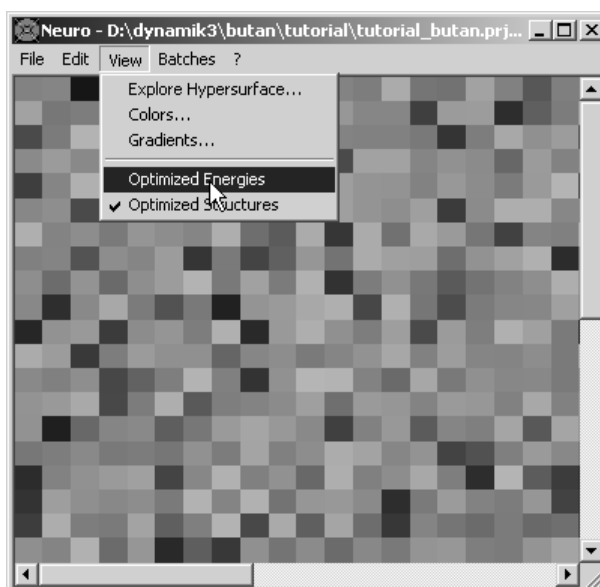
Nr. of <i>epochs</i>	50	Anzahl der Trainingsdurchläufe mit je einem kompletten Datensatz
delta	0.02	Ursprünglich ein Parameter der maximalen Distanz $d$ nach Gleichung (2.7), bei der eine Struktur noch als Fit für ein Neuron angesehen wird. Aufgrund von Normalisierungsproblemen inzwischen ohne Funktion.
maxlearnDist ( $\Delta_{\max}(1)$ )	10	Maximaler Lernradius $\Delta_{\max}(1)$ nach Gleichung (2.15) in Neuronen.
maxlearnVal ( $f_0$ )	0.15	Maximaler Adaptionfaktors $f_0$ aus Gleichung (2.12)
neighborhood function	parabolisch	Funktion der Abstandsabhängigkeit des Adaptionfaktors $f$ , Gleichung (2.18), angegeben als Wertetabelle für $\vartheta(\zeta)$ im Bereich $0 \leq \zeta \leq 1.0$ .
Fixed nr of <i>epochs</i>	<i>true</i>	Definiert, ob der Parameter „Nr of <i>Epochs</i> “ berücksichtigt werden soll, oder ob eine Heuristik den Trainingsfortgang beurteilen und den Prozess terminieren soll. Empfohlen ist aus Gründen der Reproduzierbarkeit der Wert <i>true</i> .
Nonlinear (quadratic) learning decay	<i>false</i>	<i>True</i> : verwende Gleichung (2.17) b) <i>False</i> : verwende Gleichung (2.17) a)
Use net with lowest distance sum	<i>false</i>	Verwendet eine Heuristik zur Optimierung der Qualität des erhaltenen Netzes. Die Heuristik ist nicht ausreichend getestet für regulären Einsatz.

In Bezug auf die Rechenzeit zu beachten ist, dass das Training des Netzes linear mit Anzahl der Neuronen, der Anzahl der *epochs* sowie der Anzahl der Trainingsparameter skaliert. Im Falle des Trainings von Diederwinkeln aus Ketten oder linearen *backbones* wächst die verwendete Rechenzeit damit nur linear in Bezug auf die Molekülgröße, also  $O(N)$  mit Anzahl der Atome  $N$ . Dies bedeutet, dass das Training des Netzes rechenzeitmäßig bei größeren Molekülen immer kleinere Anteile der Rechenzeit ausmacht (siehe Diskussion des MINIX-Moduls).

#### COLORIX:

Das Modul COLORIX ordnet den Neuronen die Energie der nicht geometrieoptimierten Prototypstrukturen zu. Rechenzeitmäßig ist dieses Modul prak-

tisch zu vernachlässigen und skaliert konstant  $O(1)$  mit der Anzahl der Atome im Molekül. Das Starten erfolgt analog zu Abbildung 36. Die erhaltene Energiehyperfläche kann durch Anzeige der nicht optimierten Energien als Option im Menü „View/Optimized Energies“ angezeigt werden, siehe Abbildung 38.

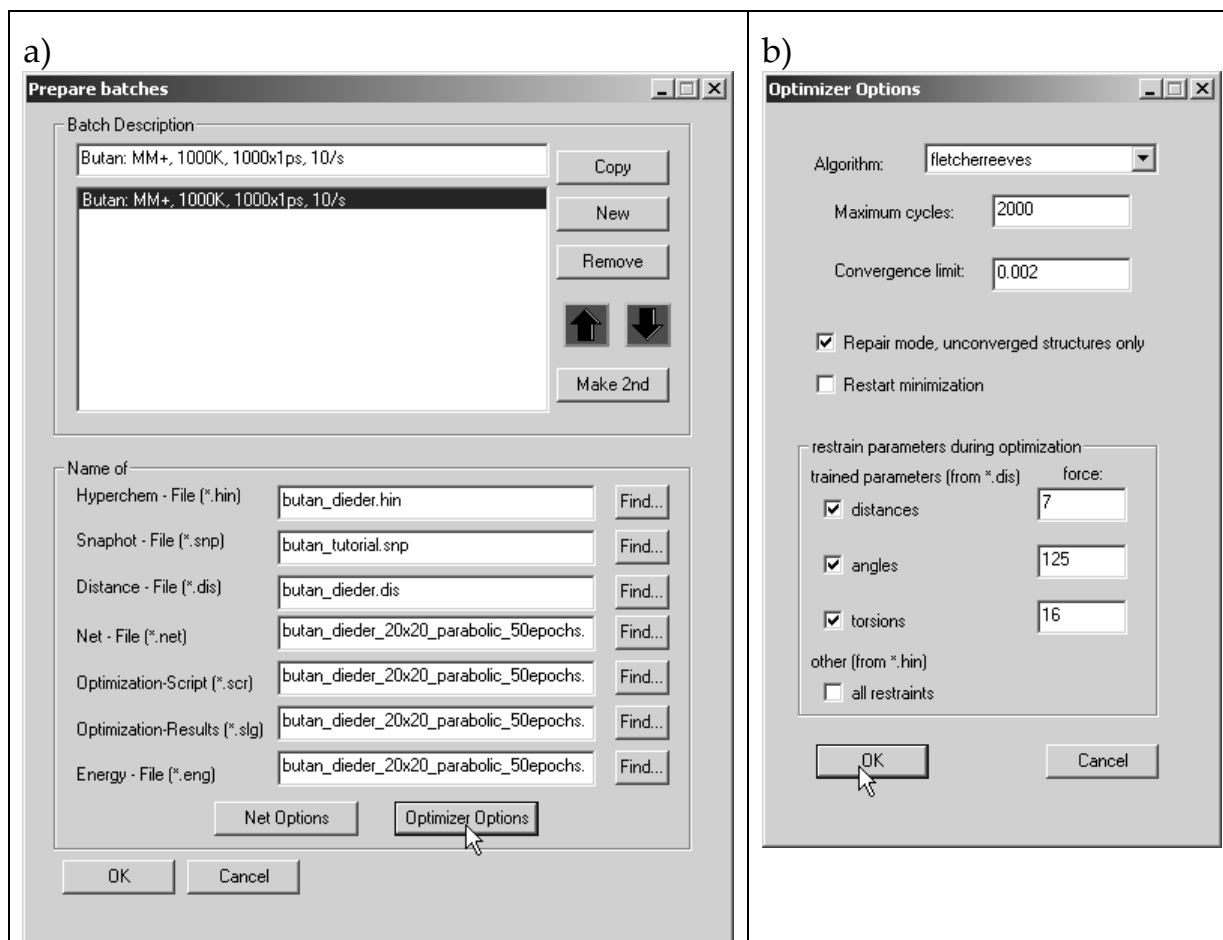


**Abbildung 38:** Tutorial COLORIX, Anzeige der nicht optimierten Energiehyperfläche

Die obige Abbildung illustriert die fehlende Aussagekraft dieses Verfahrens zumindest im vorliegenden Fall. Die innere Energie des Butan-Moleküls bei 1000K übersteigt um Größenordnungen die konformellen Barrieren der MM+ Butan-Hyperfläche aus Abbildung 29. Die Netzoberfläche zeigt nur „thermisches Rauschen“ an.

MINIX1:

Das MINIX1 Modul minimiert die Energie der Prototypstrukturen aus dem PROTOTYPIX Modul unter Verwendung des rückprojizierten Gewichtsvektors  $\mathbf{W}_{j,\text{winner}}$ , Gleichung (2.10). Es sind vor Start des Moduls die „*restrain parameters during optimization*“ im „Optimizer Options“ des Menüpunktes „Batches/Prepare Batches“ vorzunehmen. Zu empfehlen ist meist die Verwendung des HYPERCHEM-Optimieralgorithmus „Fletcher-Reeves“, siehe Abbildung 39b).

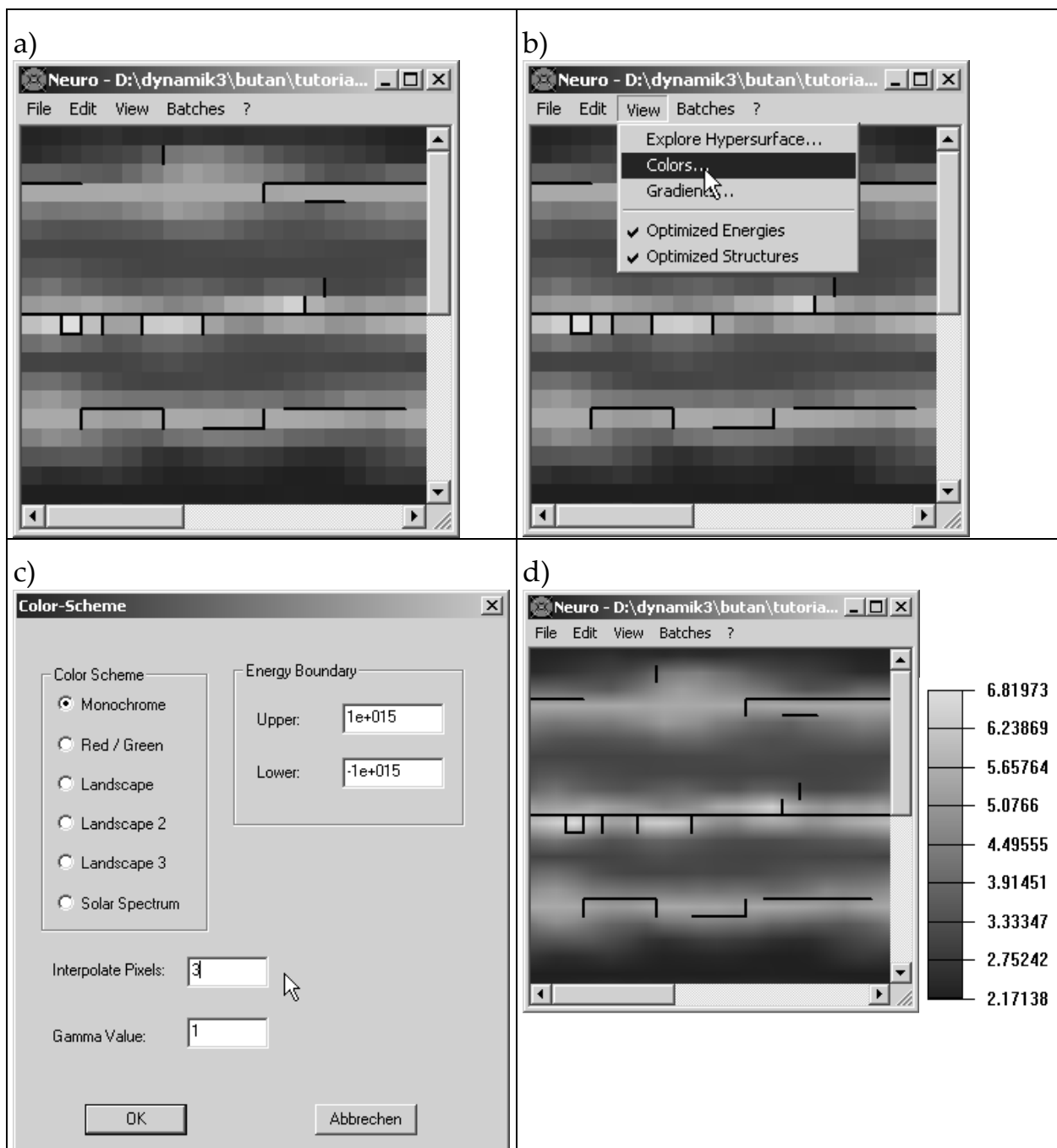


**Abbildung 39: Tutorial MINIX1, Wahl der *Optimizer Options***

Die Minimierung ist nach Durchführung der MD-Simulation einer der geschwindigkeitsbestimmenden Schritte. Je nach verwendeter Methodik (Kraftfeld, Semiempirik, ab initio) skaliert die Optimierung  $O(N^2)$  oder schlechter mit der Zahl der Atome im Molekül. Im Vergleich zur gängigen Praxis der Minimierung von Strukturen in zyklischen Abständen der Trajektorie müssen allerdings durch die Vorselektion durch das Kohonennetz meist deutlich weniger Strukturen minimiert werden.

#### COLORIX1:

Das COLORIX1 Modul ist rechenzeitmäßig gesehen wiederum zu vernachlässigen. Es wird den Netzneuronen die Energie ihrer optimierten Prototypstrukturen zugeordnet. Auswahl der optimierten Energien im „View“ Menü zeigt eine „entrauschte“ Energiehyperfläche des Butans an, erhalten aus der MD-Simulation. Subjektiv kann die Oberfläche vereinfacht werden durch Wahl eines Interpolationsalgorithmus im „View/Colors“ Menü. Dafür wird im Feld „Interpolate Pixels“ der Wert „3“ eingetragen, was die weicheren Farbverläufe in Abbildung 40 d) bewirkt. Grundsätzlich zeigen alle Netzkarten in Abbildung 40 jedoch identische Informationen.

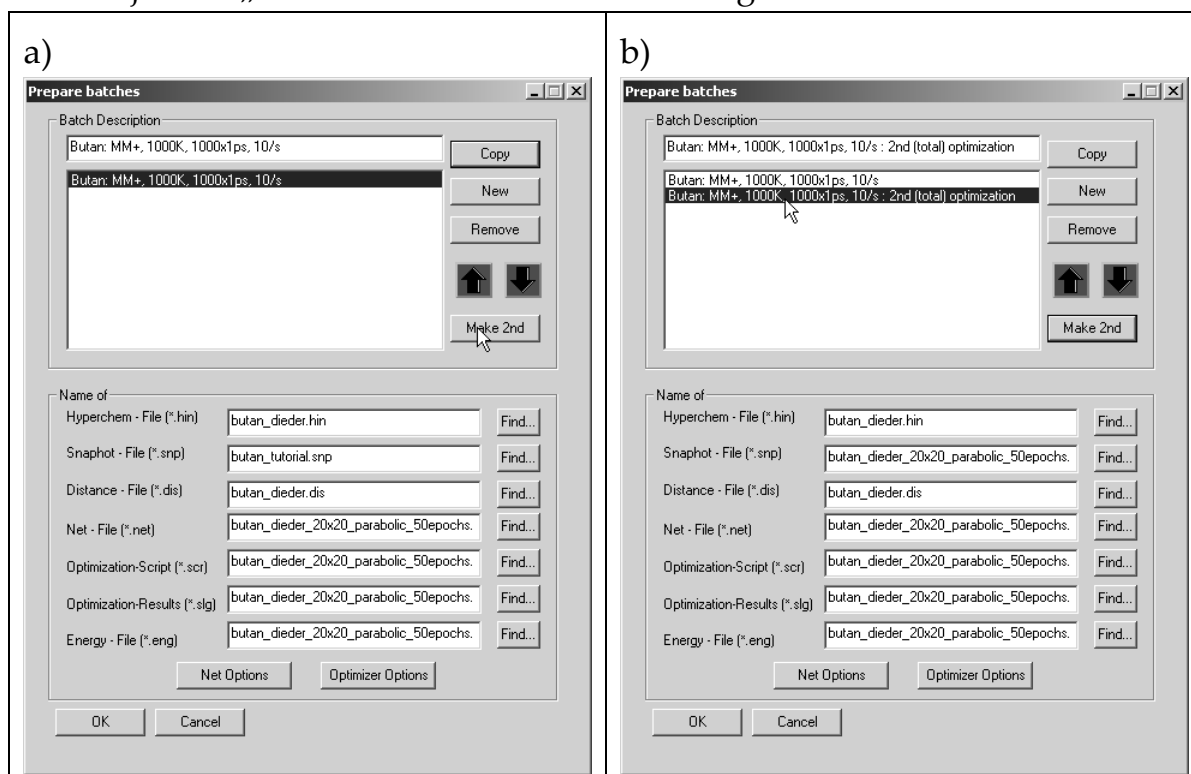


**Abbildung 40: Tutorial COLORIX1, Wahl eines Interpolationsalgorithmus für die Darstellung der Netzoberfläche**  
**a) nicht interpoliert**  
**d) interpoliert, mit Energieskala in kcal/mol**

MINIX2, COLORIX2:

Die Module MINIX2 und COLORIX2 sind von der Implementierung her Sonderfälle, da keine eigenen Routinen erstellt wurden, sondern die MINIX1/COLORIX1 Routinen auf die bereits mit *restraints* optimierten Prototypenstrukturen angewandt werden. Es wird ein neuer Batcheintrag mit dem generischen Namenszusatz *.2nd* zu allen neu zu erstellenden Dateien erzeugt und für die Durchführung der Rechnungen sowie die Anzeige der Ergebnisse verwendet. Dazu ist der Aufruf des Programms MAKE\_2ND (Kapitel 5.3.12)

erforderlich. Eine Möglichkeit besteht in der Wahl des „*Make 2nd*“-Buttons im Menü „*Batches/Prepare Batches*“, siehe Abbildung 41. Das MINIX2/COLORIX2 Modul kann dann als MINIX1/COLORIX1 Modul des generierten neuen Batcheintrags gestartet werden. Eine andere Variante ist das Starten der MINIX2 und COLORIX2 Module mit dem ursprünglichen Batcheintrag, welcher jedoch „hinter den Kulissen“ exakt das gleiche bewirkt.



**Abbildung 41: Tutorial MINIX2, Vorbereitung von MINIX2 und COLORIX2**

Nach Durchführung des MINIX2, COLORIX2 Moduls wird eine neue Hyperfläche mit den vollständig, also *restraint*freien optimierten Energien erhalten. Da die erhaltenen Netzgradienten (Gleichung (2.32)) nach Optimierung meist eine sehr viel klarere Strukturierung der Hyperfläche erlauben als vorher, werden sie per Voreinstellung verwendet. Dies lässt sich über das Menü „*View/Gradients*“ jedoch umstellen. Die erhaltenen Hyperflächen sind in Abbildung 42 und Abbildung 43 dargestellt.

Sehr schön ist in Abbildung 43 zu erkennen, dass nach der Optimierung mit *restraints* nicht nur die Minima im Zentrum der Cluster, sondern auch die Übergangszustände an den Clustergrenzen repräsentiert werden, während nach der vollständigen Optimierung natürlich nur die Minima verbleiben. Zu diesem Zweck sind *Newman*-Projektionen des Butans auf der Netzkarte eingefügt.

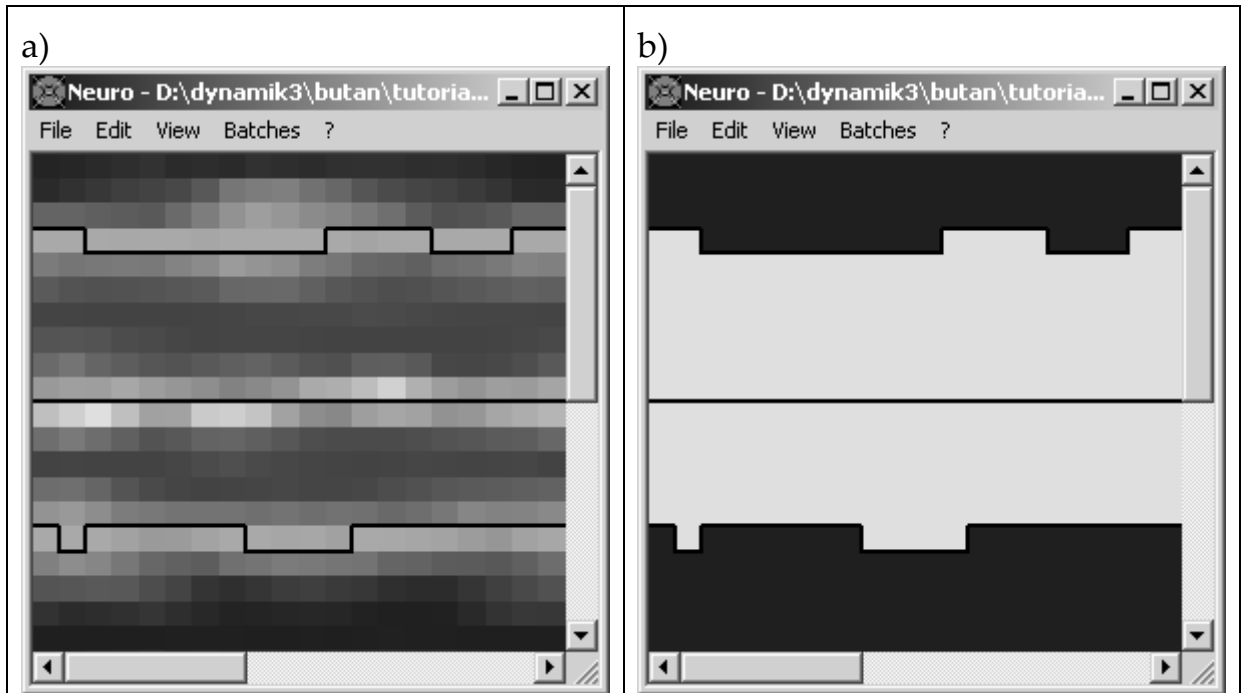


Abbildung 42: Tutorial MINIX2, Kohonen-Hyperfläche des Butans  
 a) *restrained* (View/Optimized Energies) aus  
 b) *unrestrained* (View/Optimized Energies) an

Im Vergleich zu Abbildung 40a) weist die Hyperfläche nun eine klare Strukturierung auf.

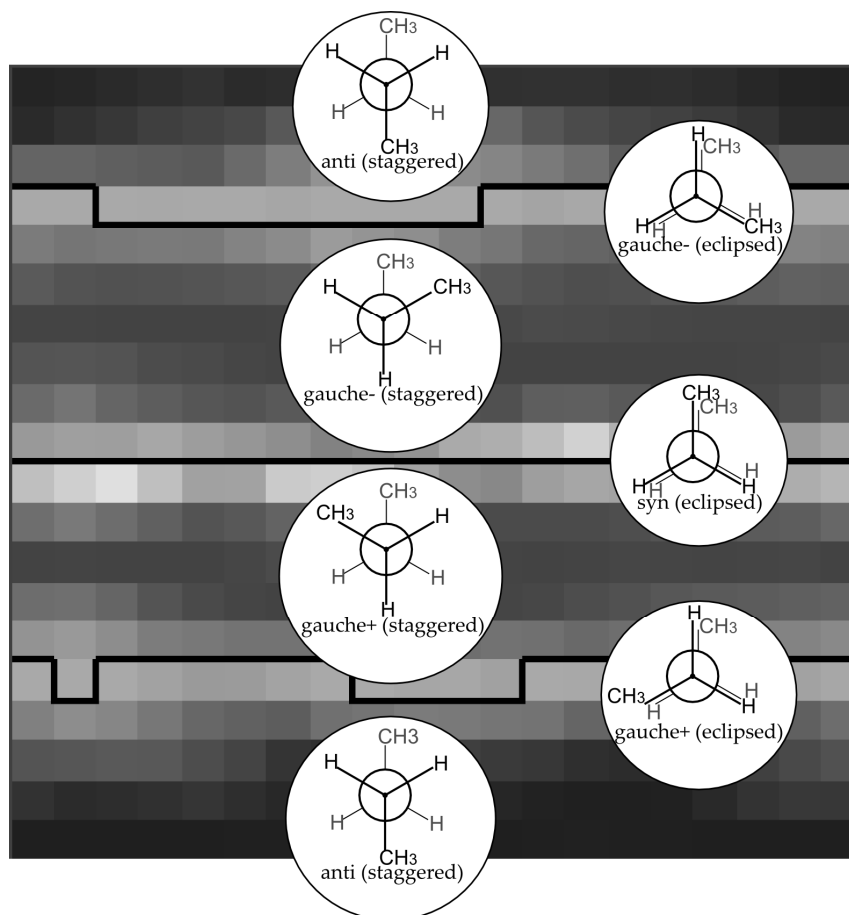
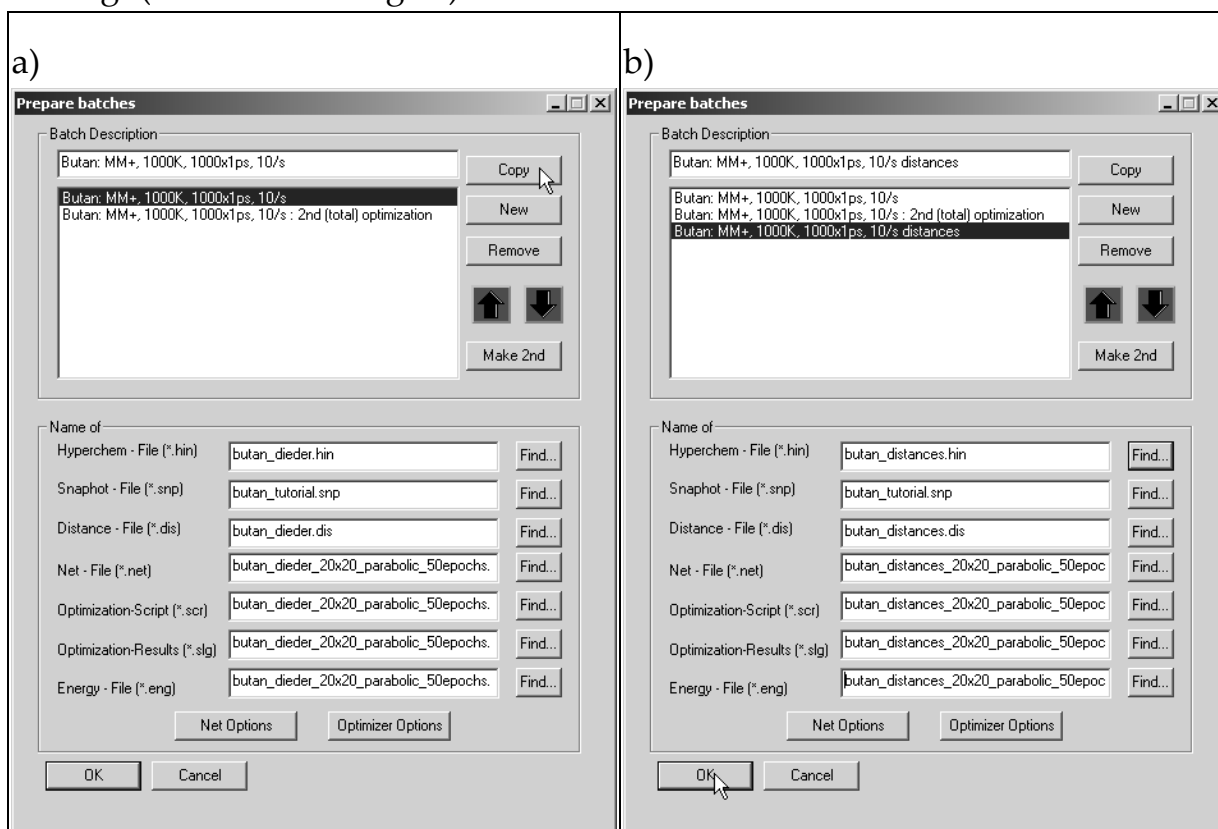


Abbildung 43: Zuordnung der Cluster auf der *restrained* Kohonen-Hyperfläche, links die Minima, rechts die Übergangszustände.

### 2.8.1.3 Tutorial: Training im Batchbetrieb

Anhand der Kohonenkarte für die vollständige Schweratom-Distanzmatrix als Trainingsparameter soll im Folgenden das automatisierte Training des Netzes für obige MD-Simulation gezeigt werden. Zuerst wird eine HYPERCHEM Datei mit den zu trainierenden Distanzen als *selections* nach Abbildung 33 sowie Anwendung des HYPERCHEM Skripts MAKE\_DISTANCE\_MATRIX.TK, Kapitel (5.2.5) erstellt unter dem Namen butan\_distances.hin gespeichert. Des Weiteren wird an dieser Stelle ein passender Batcheintrag durch Kopie und Modifikation des vorhandenen Batcheintrags für das Training der Diederwinkel erzeugt (siehe Abbildung 44).



**Abbildung 44: Tutorial automatisiertes Training, Erstellung eines Batcheintrags mittels Copy-Button**

Das sequentielle Durchführen aller Module kann durch Anwahl aller Module im Menü „Batches/Start Batches“ erreicht werden, siehe Abbildung 45.



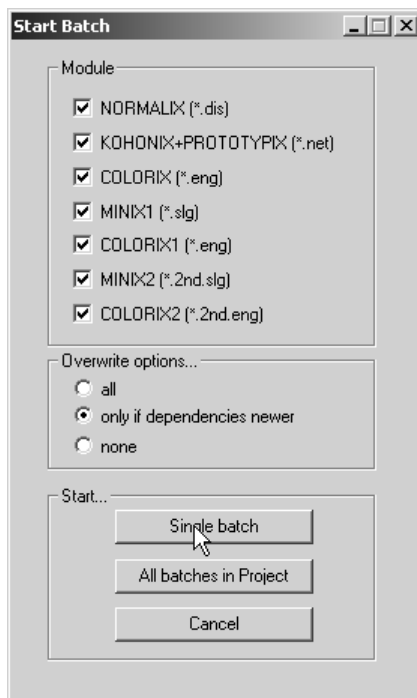


Abbildung 45: Tutorial automatisiertes Training, Start der Module

Die erhaltenen Hyperflächen sind in Abbildung 46 aufgetragen.

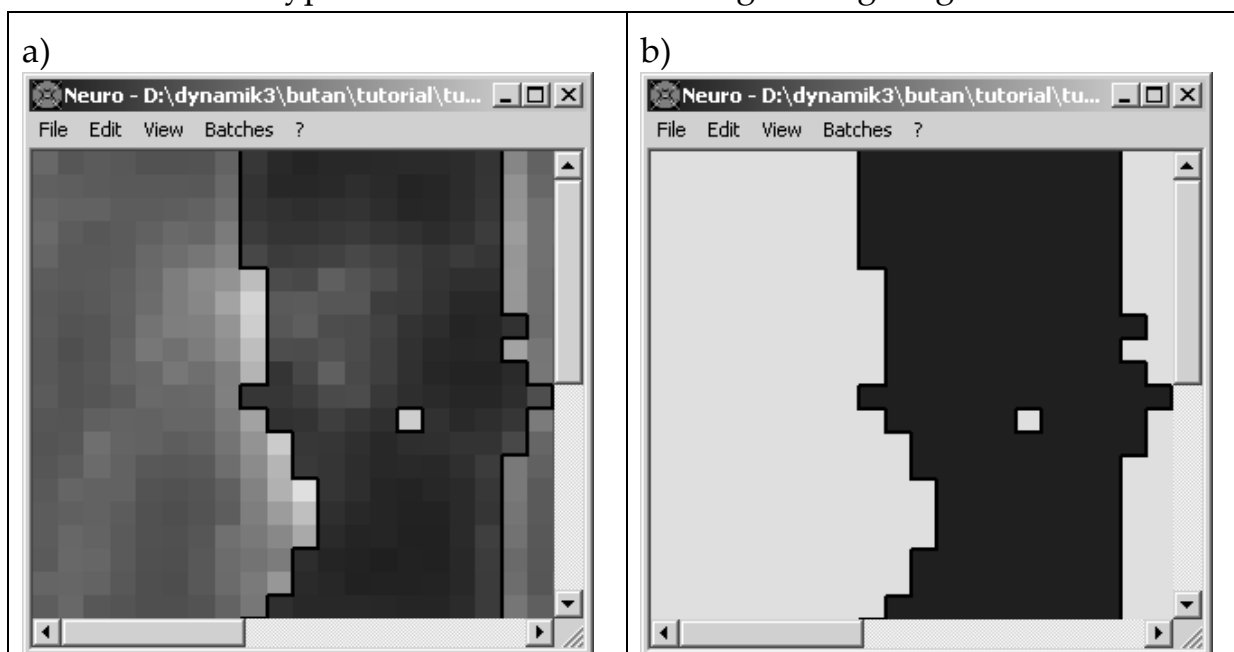


Abbildung 46: Tutorial automatisiertes Training, Kohonen-Hyperfläche des Butans nach Training der Schweratom-Distanzmatrix  
 a) *restrained* optimierte Energien  
 b) *unrestrained* optimierte Energien

Deutlich zu erkennen ist, dass die Netzoberfläche sehr viel unruhiger geworden ist und ein Ausreißer in *gauche*-Konformation sich in die Fläche der *anti*-Konformation „verirrt hat“. Außerdem gibt es keine Trennung der *gauche*- und *gauche*<sup>+</sup>-Konformationen. Dies erklärt sich aus der Indifferenz der Distanzmatrix gegenüber Enantiomerie im Molekül. Es ist bereits in diesem

einfachen Beispiel zu erkennen, dass der normalerweise nichttrivialen Auswahl der zu trainierenden internen Koordinaten große Bedeutung zukommt.

### 2.8.1.4 Tutorial: Hypersurface Explorer

Das folgende Tutorial behandelt die Projektion der Eingangsdaten als Trajektorie auf die Netzoberfläche. Dies ist für kontinuierliche Daten, z.B. MD- oder MC-Simulationen sinnvoll, um den Verlauf der Trajektorie zu verfolgen. Aber auch für diskontinuierliche Strukturensamples eröffnet dieses Analyseverfahren die Möglichkeit, verschiedene Datensätze auf Überlappungen im repräsentierten Konformationsraum zu überprüfen. Dies ist mit den herkömmlichen Methoden zur Analyse hochdimensionaler Datensätze aus Kapitel 1.3, wenn überhaupt, dann nur mit sehr viel mehr manuellem Aufwand erreichbar. Als Netzoberfläche dient die Oberfläche aus Abbildung 42 b). Nach Laden der Projektdatei und Auswahl des „.2nd“-Batches, wird nun der Menüpunkt „View/Explore Hypersurface“ angewählt, welcher über den „Hypersurface Explorer“ die Module (TRAJECTORIX, EXTRACTIX und PROJECTIX zugänglich macht, siehe Abbildung 47:

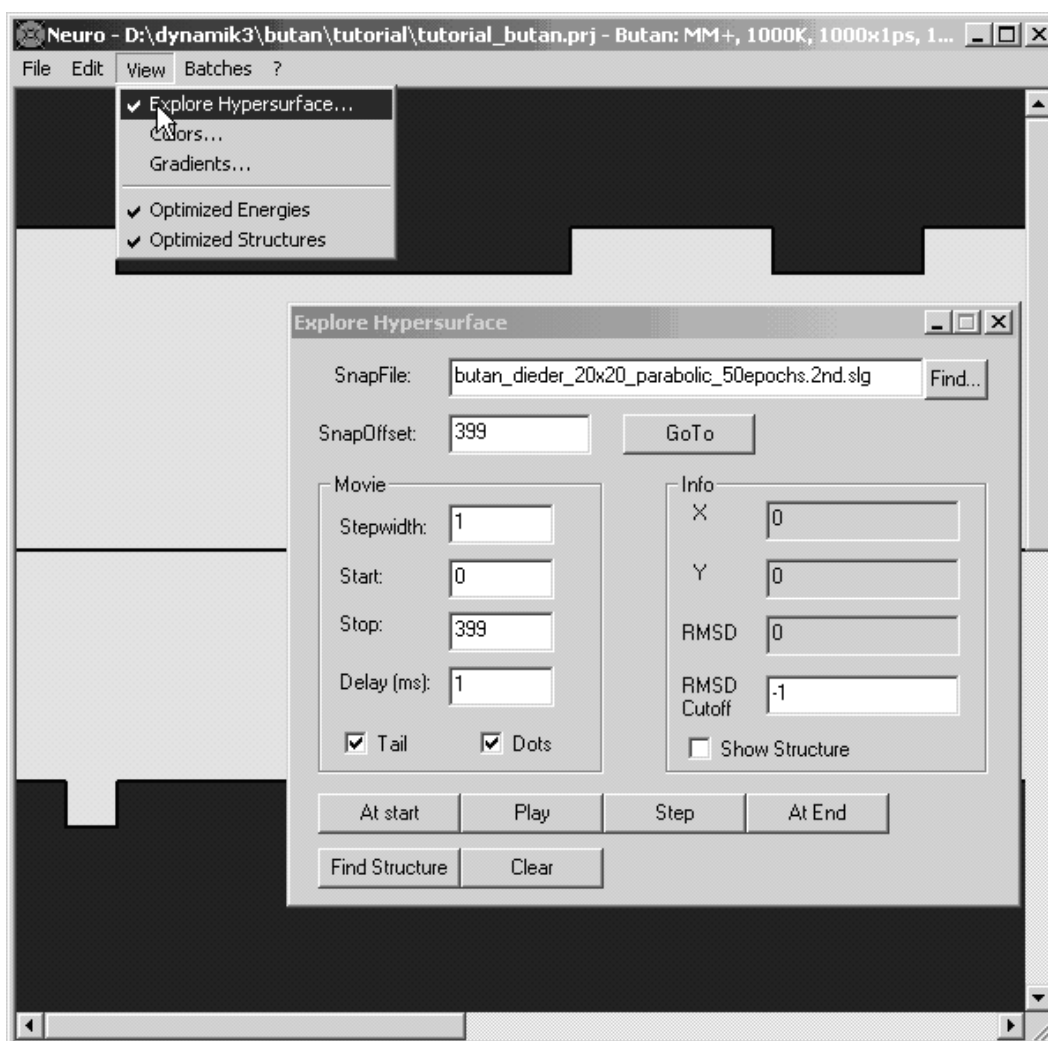


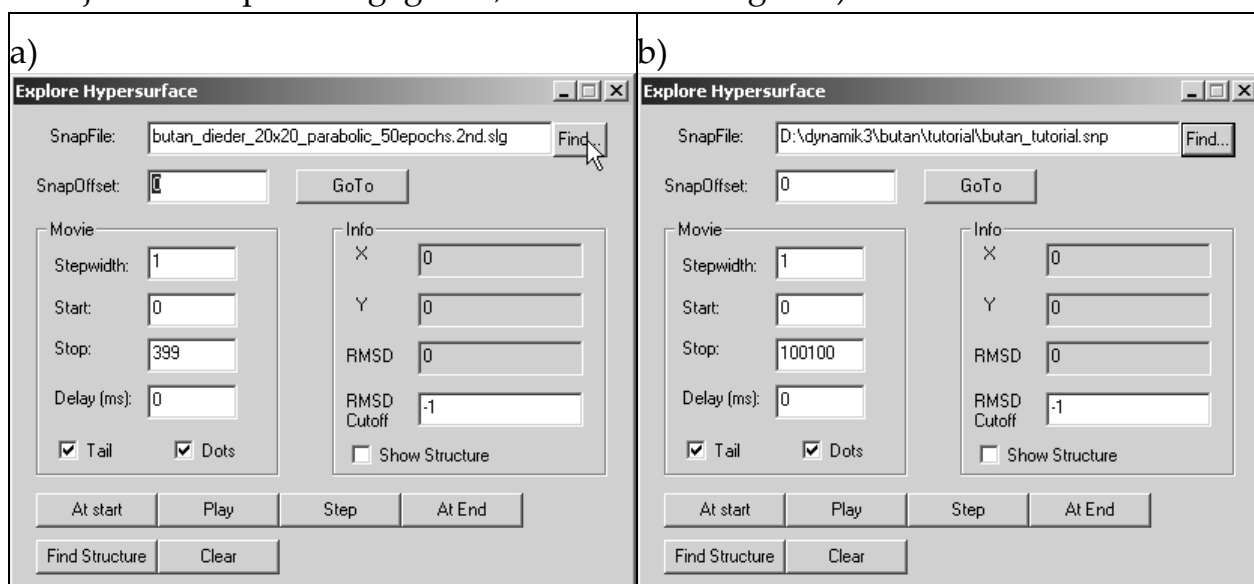
Abbildung 47: Tutorial: Hypersurface Explorer

Verwendet wird in diesem Fall die Netzoberfläche der zweiten, *restraint*freien Optimierung mit dem Diederwinkel als Trainingsatz, da diese eine übersichtlichere Gliederung durch die Netzgradienten (siehe Gleichung (2.32)) erbringt. Als Voreinstellung wird in der Rubrik *Snapfile* je nach Einstellung im *View*-Menü eines der Strukturensamples des „*Prepare Batch*“-Dialogs aus Abbildung 35 angezeigt:

„View/Optimized Energies“ an → „Optimization Results (\*.slg)“

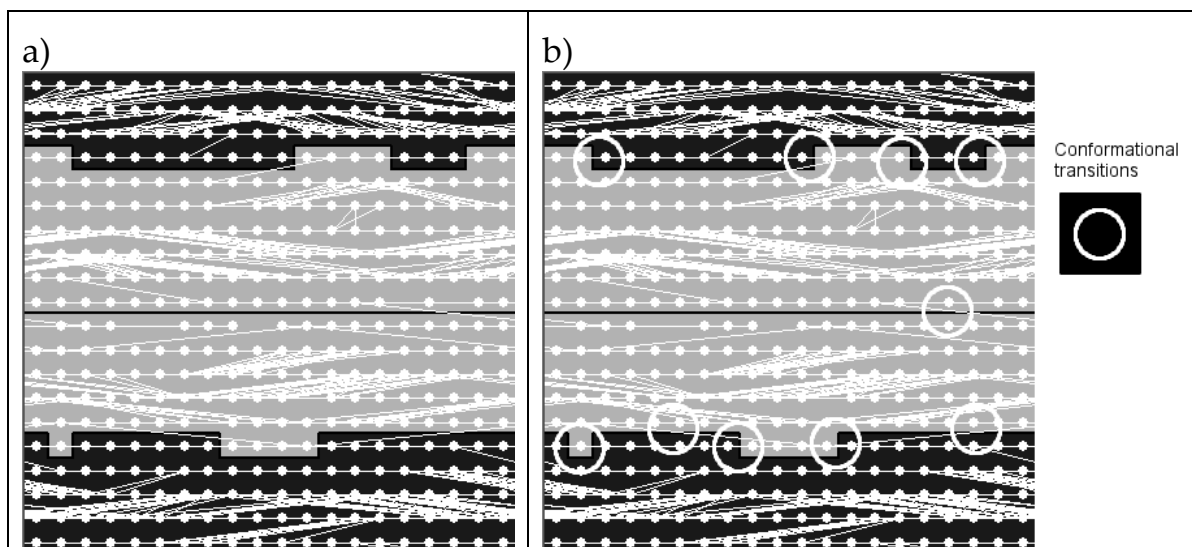
„View/Optimized Energies“ aus → „Snapshot File (\*.snp)“

Da das „*Snapshot File*“ des zweiten Optimierdurchgangs aus den optimierten Strukturen des ersten Durchgangs besteht, ist die Voreinstellung an dieser Stelle in der Regel unbrauchbar. Es wird stattdessen die ursprüngliche MD-Trajektorie explizit angegeben, siehe Abbildung 48 a).



**Abbildung 48:** Tutorial Hypersurface Explorer, Wahl der Trajektorie

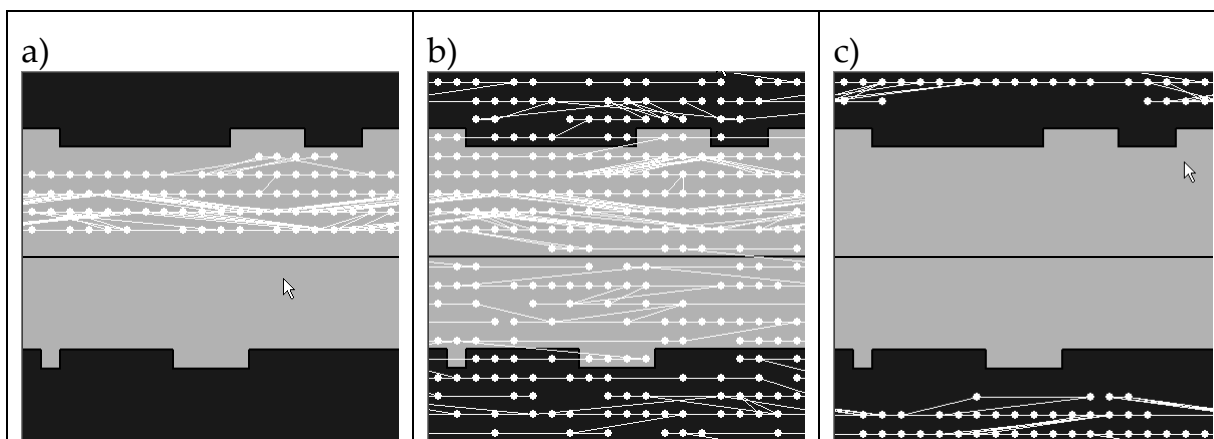
Mit den Einstellungen aus Abbildung 48 b) wird die komplette Trajektorie auf die Netzoberfläche projiziert von der ersten ( $Start=0$ ) bis zur letzten ( $Stop=100100$ ) Struktur. Besuchte Strukturen werden mit einem weißen Punkt auf der Netzoberfläche eingetragen (*Dots=enabled*). Konsecutive (in der MD-Simulation aufeinanderfolgende) Strukturen werden mit einem weißen Strich verbunden (*Tail=enabled*). Die erhaltene Trajektorie ist in Abbildung 49a) dargestellt.



**Abbildung 49:** Tutorial Hypersurface Explorer, a) komplette Trajektorie, b) Kreise zeigen Konformationsübergänge an

Aus der Trajektorie können Rückschlüsse über die Topologie der abgedeckten Hyperfläche sowie den Verlauf der MD-Simulation abgeleitet werden. So ist in obigem Bild zu erkennen, dass das Butan-Molekül der Simulation sich meist eine ganze Weile innerhalb einer Konformation aufhält, bevor es in eine benachbarte wechselt. Dies kann in diesem Spezialfall einer eindimensionalen Hyperfläche aus der Dominanz der horizontalen Verbindungslinien geschlossen werden. Das Überwinden einer Konformationsbarriere wird übersichtlich als das Überschreiten einer starken Gradientenlinie visualisiert, siehe Abbildung 49b.

Das Auftragen von Ausschnitten der Trajektorie ist eine weitere Möglichkeit, Informationen über die Moleküldynamik zu gewinnen, siehe Abbildung 50. In Ausschnitt a) verbleibt die MD-Simulation komplett in einem *gauche*-Minimum des Butans. Ausschnitt b) enthält sämtliche Konfigurationen, in Ausschnitt c) wiederum ist die Simulation im globalen Minimum (*anti*) gefangen.



**Abbildung 50:** Tutorial Hypersurface Explorer – Intervalle  
 a) Strukturen 0 ... 1000 (von 0 bis 10 ps der MD-Simulation)  
 b) Strukturen 1000 ... 2000 (von 10 bis 20 ps der MD-Simulation)  
 c) Strukturen 10000 ... 11000 (von 100 bis 110 fs der MD-Simulation)

### 2.8.1.5 Training der Grid-Hyperfläche

Als Haupteinflussgrößen für die Qualität der erhaltenen Netzkarten werden aufgrund von Vorversuchen die in Tabelle 14 aufgeführten Variablen angesetzt.

**Tabelle 14:** untersuchte Einflussgrößen für die Qualität der Netzkarten

Parameter	Erläuterung
maxlearnDist ( $\Delta_{\max}(1)$ )	Maximaler Lernradius $\Delta_{\max}(1)$ nach Gleichung (2.15) in Neuronen.
maxlearnVal ( $f_0$ )	Maximaler Adaptionfaktor $f_0$ aus Gleichung (2.12).
neighborhood function	Funktion der Abstandsabhängigkeit des Adaptionfaktors $f$ , Gleichung (2.18), angegeben als Wertetabelle für $\vartheta(\zeta)$ im Bereich $0 \leq \zeta \leq 1.0$ .
Nr of epochs	Anzahl der Trainingsdurchläufe mit je einem kompletten Datensatz.
NETX, NETY	Netzgröße, hier quadratisch, NETX=NETY.
quadratic learning decay	Verwendete Zeitfunktion nach Gleichung (2.17).

Ebenfalls von großem Einfluss für die subjektive Qualität der erhaltenen Netzkarten sind selbstverständlich die Qualität der für das Training verwendeten Strukturensamples sowie die Auswahl der zum Training verwendeten internen Parameter. Die Wahl kleiner Moleküle als Benchmark hat als Vorteil, dass die Hyperfläche und die optimale Auswahl an trainierten internen Parametern bekannt und somit die Ergebnisse leicht verifizierbar sind. Es werden zunächst mit dem Skript GRID.TCL (siehe Kapitel 5.2.2) durch Variation des zentralen Diederwinkels in  $1^\circ$ -Schritten die Strukturen aus Abbildung 29 erzeugt. Die Verwendung des CC-Diederwinkels als trainierter interner

Koordinate entspricht der traditionellen Darstellung der Butan-Hyperfläche, siehe Abbildung 29.

In Tabelle 15 bis Tabelle 20 sind die Ergebnisse für die in Kapitel 2.5 aufgeführten Qualitätsparameter bei der Variation jeweils eines einzigen Netzparameters angegeben. Alle nicht variierten Netzparameter haben folgende Werte: [ $f_0=0.15$ ,  $epochs=1000$ ,  $neighborhood\ function = \text{parabolisch}$ ,  $NETX=NETY=21$ ,  $linear\ learning\ decay$ ]. Für die übrigen Netzparameter wurden in Vorversuchen ermittelte robuste, aber wahrscheinlich nicht optimale Werte gewählt. Es wurden jeweils 10 Trainingsdurchläufe durchgeführt, angegeben sind:

- Deskriptor: Mittelwert des Deskriptors.
- $\sigma$ : Standardabweichung aus einer Stichprobe.  

$$\sigma_x = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2}$$
 Dieser Wert wird in den Graphen als Fehlerbalken angezeigt.
- $relerr$ : Der relative Fehler des Deskriptors.  

$$relerr\% = \frac{\sigma_x}{\bar{X}_x} 100\%$$
- $min$ ,  $max$ : minimale und maximale Werte des Deskriptors über alle Messpunkte zur Überprüfung auf „Ausreißer“
- $ndata$ : Anzahl der Messpunkte (Trainingsdurchgänge)

Für die Ermittlung der RMSDCF-Werte wird eine Klassifizierung der Butan-konformationen mittels des Programms CLASSIFY\_BUTAN\_SNP.EXE durchgeführt.

#### **2.8.1.5.1 Variation der epochs**

Die folgende Tabelle 15 enthält die statistischen Daten der Qualitätsparameter RMSDCF, RMSG und RMSDSN, welche bei mehrfachen Netztrainingsläufen mit den oben angegebenen Trainingsparametern unter Variation der Anzahl der *epochs* erhalten werden.

Tabelle 15: Butan - Variation der *epochs*

<i>epochs</i>	Ø RMSDCF (2.24)	$\sigma$	relerr%	min	max	ndata
2000	0.00705	0.00261	36.95	0.00257	0.01050	10
1000	0.00665	0.00286	42.96	0.00113	0.01028	10
500	0.00687	0.00330	48.07	0.00345	0.01255	10
300	0.00966	0.00553	57.23	0.00298	0.01996	10
200	0.00647	0.00396	61.24	0.00226	0.01267	10
100	0.00616	0.00315	51.06	0.00264	0.01244	10
50	0.00985	0.00431	43.72	0.00416	0.01708	10
10	0.00881	0.00764	86.69	0.00334	0.02856	10
<i>epochs</i>	Ø RMSG (2.23)	$\sigma$	relerr%	min	max	ndata
2000	0.07242	0.00093	1.29	0.07070	0.07427	10
1000	0.07242	0.00060	0.83	0.07145	0.07326	10
500	0.07196	0.00065	0.90	0.07090	0.07267	10
300	0.07184	0.00079	1.10	0.07057	0.07268	10
200	0.07227	0.00056	0.77	0.07138	0.07303	10
100	0.07185	0.00064	0.89	0.07066	0.07290	10
50	0.07078	0.00067	0.95	0.06943	0.07185	10
10	0.07718	0.01554	20.14	0.06924	0.10721	10
<i>epochs</i>	Ø RMSDSN(2.25)	$\sigma$	relerr%	min	max	ndata
2000	0.00122	0.00008	6.58	0.00111	0.00137	10
1000	0.00137	0.00004	3.00	0.00133	0.00148	10
500	0.00156	0.00012	7.65	0.00141	0.00179	10
300	0.00184	0.00022	11.78	0.00149	0.00215	10
200	0.00186	0.00018	9.94	0.00150	0.00207	10
100	0.00204	0.00020	9.66	0.00182	0.00243	10
50	0.00250	0.00034	13.75	0.00189	0.00294	10
10	0.00222	0.00054	24.38	0.00141	0.00325	10

Grafisch aufgetragen werden die Ergebnisse für alle drei Deskriptoren inklusive Fehlerbalken in Abbildung 51. Die Anzahl der *epochs* ist im eigentlichen Sinne keine optimierbare Größe, gesucht wird eine robuste untere Grenze. Der RMSDCF-Wert weist eine hohe Streuung auf.

Eine geringe Streuung, sprich hohe Reproduzierbarkeit, weisen die RMSG- sowie die RMSDSN-Werte auf. Beide Kurven zeigen einen lokalen Extremwert im Bereich kleiner *epochs*, die asymptotische Annäherung an einen Grenzwert erfolgt erst ab einem Wert von ca. 50 *epochs*. Daraus folgt, dass die Anzahl der *epochs* in den folgenden Rechnungen definitiv über 50 liegen muss.

Ab ca. 500 *epochs* lassen sich im Rahmen der Fehlergrenzen insgesamt keine nennenswerten Verbesserungen für den RMSG-Wert mehr erkennen, was bei 360 Strukturen pro *epoch* einen MTS(RMSG)-Wert von 180,000 ergibt, siehe Gleichung (2.26). Der RMSDSN-Wert erreicht im Rahmen der Fehlergrenzen ab 2000 *epochs* einen konstanten Wert,  $MTS(RMSDSN)=720,000$ .

MTS(RMSDCF) wird aufgrund der anhaltend hohen Standardabweichung der RMSDCF-Werte nicht berechnet.

Visuell lassen sich keine Qualitätsunterschiede in den erhaltenen Karten erkennen. Im Folgenden wurden 1000 *epochs* als robuster Standardwert zum Training verwendet.

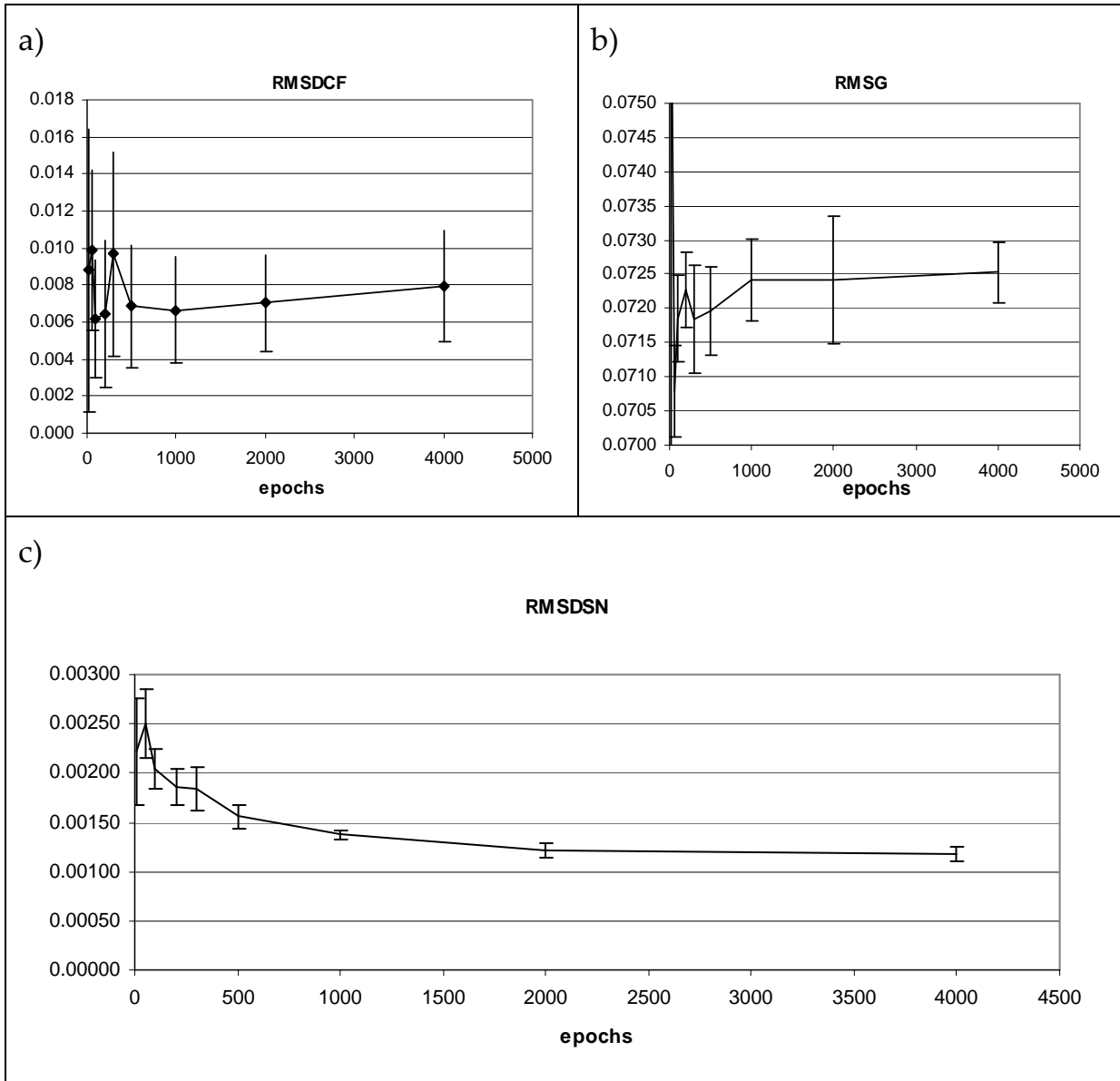


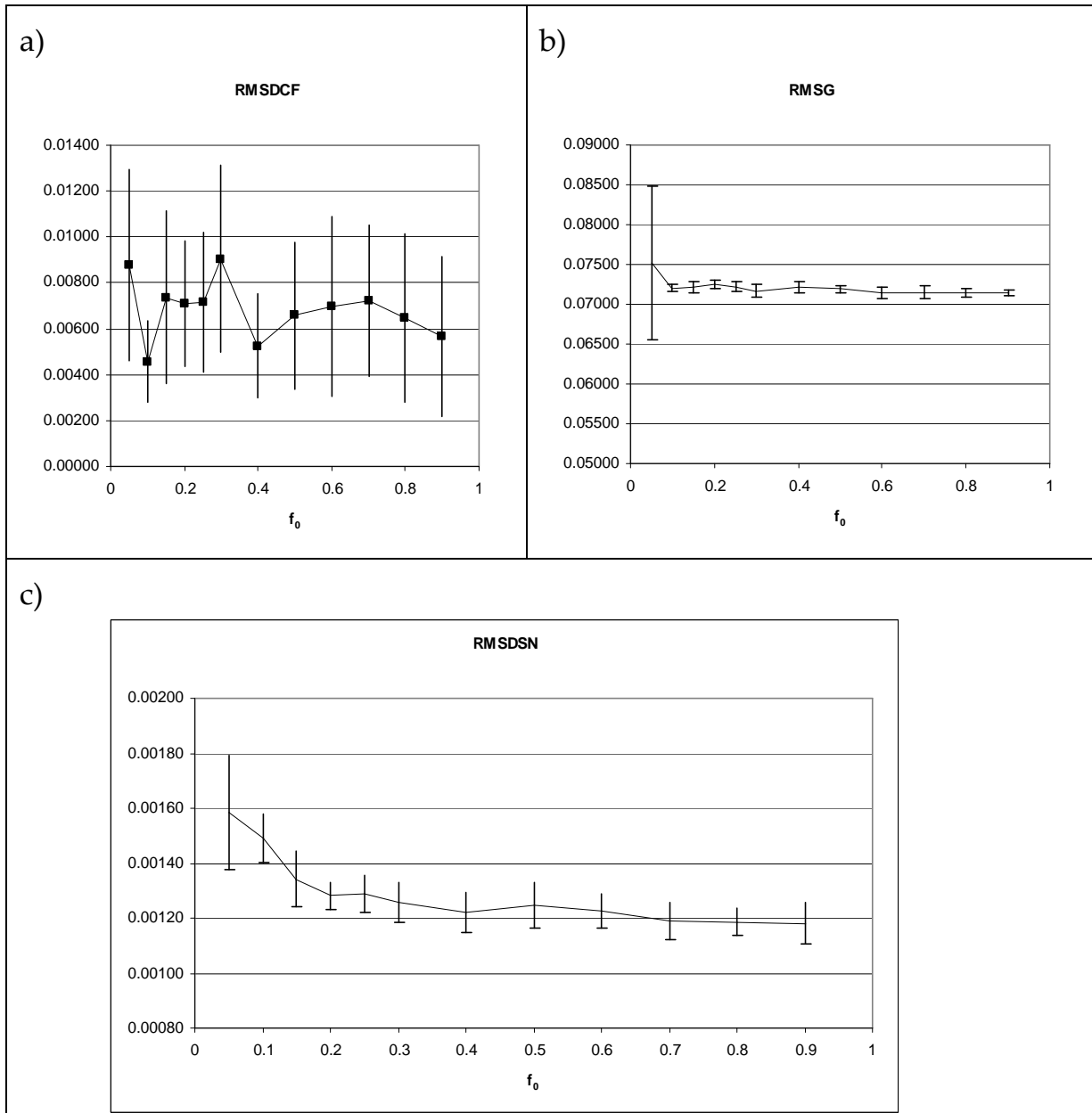
Abbildung 51: Variation der *epochs*, a) RMSDCF, b) RMSG, c) RMSDSN



2.8.1.5.2 Variation von  $f_0$ Tabelle 16: Variation des maximalen Adaptionfaktors  $f_0$  nach Gleichung (2.12)

$f_0$	$\emptyset$ RMSDCF (2.24)	$\sigma$	relerr%	min	max	ndata
0.05	0.00876	0.00417	47.63	0.00072	0.01382	10
0.1	0.00457	0.00176	38.61	0.00156	0.00702	10
0.15	0.00736	0.00376	51.08	0.00226	0.01264	10
0.2	0.00710	0.00276	38.81	0.00382	0.01189	10
0.25	0.00716	0.00303	42.34	0.00323	0.01227	10
0.3	0.00905	0.00408	45.05	0.00264	0.01464	10
0.4	0.00524	0.00227	43.25	0.00167	0.00813	10
0.5	0.00657	0.00322	48.93	0.00257	0.01046	10
0.6	0.00695	0.00392	56.35	0.00257	0.01423	10
0.7	0.00721	0.00331	45.90	0.00323	0.01382	10
0.8	0.00649	0.00366	56.45	0.00334	0.01523	10
0.9	0.00567	0.00350	61.79	0.00113	0.01264	10
$f_0$	$\emptyset$ RMSG (2.23)	$\sigma$	relerr%	min	max	ndata
0.05	0.07509	0.00965	12.84	0.07100	0.10246	10
0.1	0.07203	0.00050	0.69	0.07122	0.07255	10
0.15	0.07213	0.00065	0.90	0.07139	0.07325	10
0.2	0.07253	0.00057	0.79	0.07147	0.07341	10
0.25	0.07223	0.00056	0.77	0.07157	0.07333	10
0.3	0.07165	0.00078	1.09	0.07073	0.07295	10
0.4	0.07207	0.00071	0.98	0.07089	0.07322	10
0.5	0.07193	0.00046	0.64	0.07144	0.07299	10
0.6	0.07144	0.00070	0.97	0.06999	0.07222	10
0.7	0.07149	0.00083	1.16	0.07029	0.07275	10
0.8	0.07140	0.00049	0.68	0.07069	0.07227	10
0.9	0.07149	0.00038	0.53	0.07095	0.07221	10
$f_0$	$\emptyset$ RMSDSN(2.25)	$\sigma$	relerr%	min	max	ndata
0.05	0.00159	0.00021	13.17	0.00118	0.00186	10
0.1	0.00149	0.00009	5.90	0.00133	0.00167	10
0.15	0.00134	0.00010	7.54	0.00117	0.00148	10
0.2	0.00128	0.00005	3.97	0.00119	0.00135	10
0.25	0.00129	0.00007	5.17	0.00118	0.00139	10
0.3	0.00126	0.00007	5.94	0.00114	0.00137	10
0.4	0.00122	0.00007	5.98	0.00112	0.00132	10
0.5	0.00124	0.00008	6.67	0.00112	0.00138	10
0.6	0.00123	0.00006	5.05	0.00112	0.00131	10
0.7	0.00119	0.00007	5.62	0.00106	0.00131	10
0.8	0.00119	0.00005	4.08	0.00111	0.00126	10
0.9	0.00118	0.00007	6.21	0.00107	0.00129	10

Der maximale Adaptionfaktor  $f_0$  scheint im Rahmen der grafischen Auswertung Abbildung 52 praktisch beliebige Werte  $\geq 0.10$  annehmen zu können. Der RMSDCF-Wert ist hier aufgrund der Fehlergrenzen insgesamt wenig aussagekräftig. RMSG- und RMSDSN-Wert zeigen weitgehend konstanten Verlauf für  $0.2 \leq f_0 \leq 0.9$ .

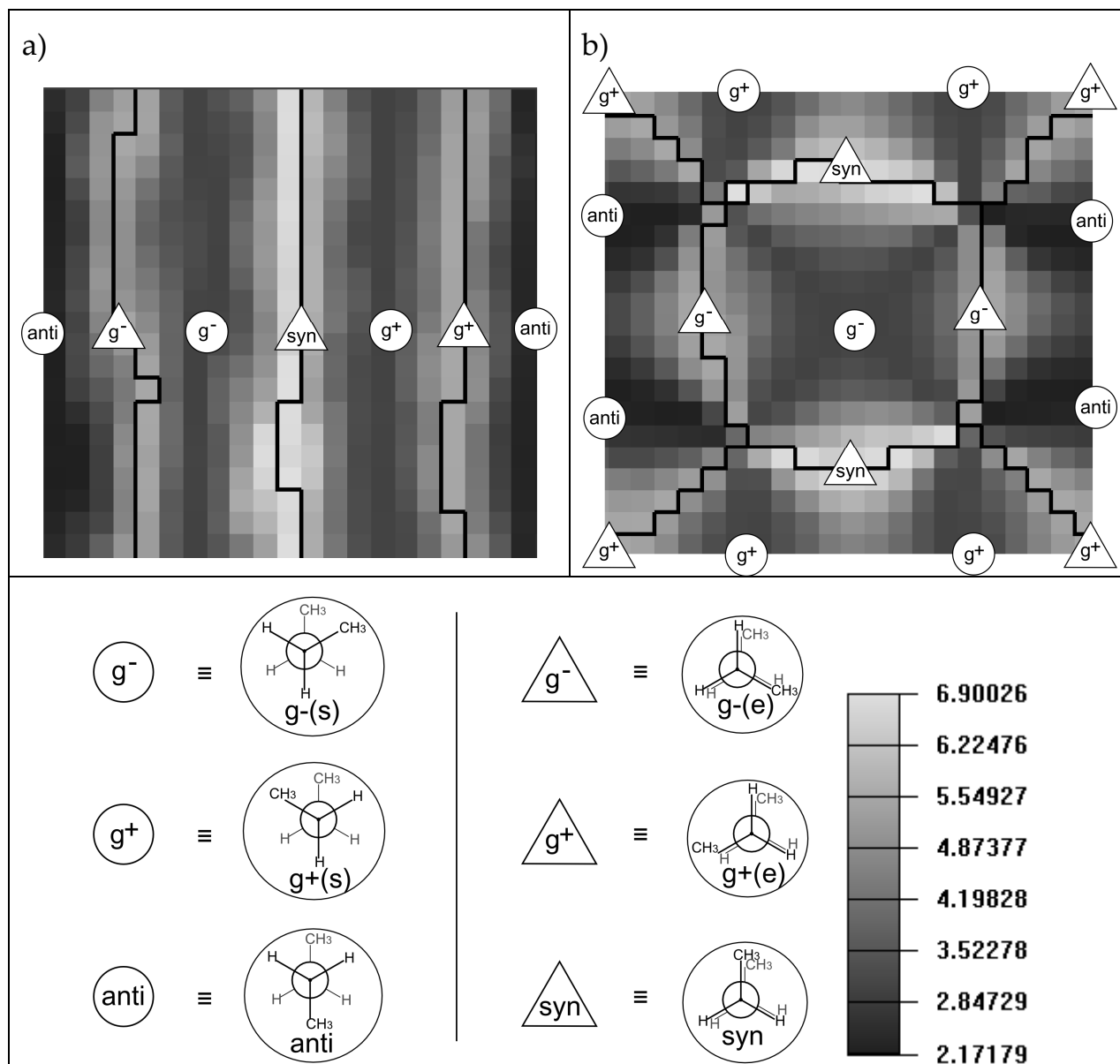


**Abbildung 52:** Variation von  $f_0$ , a) RMSDCF, b) RMSG, c) RMSDSN

Aufgrund von Vorversuchen ist bekannt, dass größere Werte von  $f_0$  ein schnelleres Training in weniger *epochs* ermöglichen, es jedoch eher zu Fehlkonformationen der Netztopologie kommen kann, welche identische Eingabedaten an unterschiedliche Stellen der Netzoberfläche projizieren und ein stabiles „lokales Minimum“ im Trainingsverlauf darstellen. Diese werden deshalb im Folgenden „degenerierte Netzarten“ genannt. Die extrem einfache „unterdimensionale“ Hyperfläche des Butans zeigt dieses Verhalten nur sehr selten und ist damit nicht besonders aussagekräftig für die Suche nach einem robusten  $f_0$ -Standardwert.

Ein Beispiel für eine solche degenerierte Netzarte (einmaliges Ereignis beim Training der 120 Netzarten aus Tabelle 16) ist in Abbildung 53 b) dargestellt. Diese Fehlkonformationen können im Vergleich zu anderen Netzarten

zuverlässig an einem hohen RMSG-Wert erkannt werden, siehe  $\max(\text{RMSG})$ , Tabelle 16.



**Abbildung 53: Kohonenkarten des Butans (Training der Grid-Hyperfläche)**  
**a) normale Netzkarte, erhalten bei  $f_0=0.05$  (RMSG=0.072)**  
**b) degenerierte Netzkarte, erhalten bei  $f_0=0.05$  (RMSG=0.102)**

In obiger Grafik sind die sowohl die Minima (als Kreise) als auch die Übergangszustände (als Dreiecke) repräsentiert und den *Newman*-Projektionen ihrer Butankonformation zugeordnet. Die Gewinnung der Übergangszustände aus den Netz-Restrains ist ein willkommener „Nebeneffekt“ des MINIX1-Moduls und zugleich ein Vorteil der NEURO-SUITE.

### 2.8.1.5.3 Variation von $\Delta_{\max}(1)$

Tabelle 17: Variation von  $\Delta_{\max}(1)$  nach Gleichung (2.15)

$\Delta_{\max}(1)$	$\emptyset$ RMSDCF (2.24)	$\sigma$	relerr%	min	max	ndata
10	0.00723	0.00344	47.51	0.00156	0.01264	10
9	0.00915	0.00333	36.41	0.00334	0.01444	10
8	0.00672	0.00380	56.51	0.00234	0.01224	10
7	0.00763	0.00410	53.79	0.00167	0.01412	10
6	0.00616	0.00365	59.36	0.00156	0.01383	10
5	0.00680	0.00432	63.53	0.00113	0.01404	10
4	0.00835	0.00379	45.37	0.00298	0.01390	10
3	0.01091	0.00601	55.12	0.00434	0.02315	10
2	0.00795	0.00401	50.38	0.00323	0.01590	10
1	0.01389	0.00985	70.92	0.00416	0.03524	10
$\Delta_{\max}(1)$	$\emptyset$ RMSG (2.23)	$\sigma$	relerr%	min	max	ndata
10	0.07220	0.00083	1.15	0.07080	0.07322	10
9	0.07254	0.00053	0.74	0.07151	0.07345	10
8	0.07226	0.00048	0.67	0.07143	0.07276	10
7	0.07167	0.00105	1.47	0.06992	0.07318	10
6	0.07239	0.00067	0.92	0.07167	0.07365	10
5	0.07255	0.00059	0.81	0.07161	0.07352	10
4	0.08454	0.01625	19.23	0.07059	0.10450	10
3	0.08766	0.01615	18.43	0.07139	0.10432	10
2	0.09713	0.01350	13.89	0.07149	0.10476	10
1	0.23772	0.01520	6.39	0.19759	0.24893	10
$\Delta_{\max}(1)$	$\emptyset$ RMSDSN(2.25)	$\sigma$	relerr%	min	max	ndata
10	0.00133	0.00009	6.41	0.00120	0.00146	10
9	0.00130	0.00006	4.31	0.00123	0.00143	10
8	0.00130	0.00009	6.98	0.00113	0.00141	10
7	0.00127	0.00009	7.46	0.00112	0.00142	10
6	0.00123	0.00007	5.59	0.00113	0.00138	10
5	0.00118	0.00008	6.66	0.00106	0.00130	10
4	0.00106	0.00012	11.35	0.00091	0.00123	10
3	0.00104	0.00016	15.31	0.00087	0.00122	10
2	0.00088	0.00009	9.69	0.00078	0.00106	10
1	0.00115	0.00004	3.83	0.00108	0.00121	10

Der Einfluss der Größe der Trainingsumgebung,  $\Delta_{\max}(1)$  scheint im Bereich 10...5 (100%...50 % der maximal möglichen Trainingsdistanz) sehr gering zu sein. RMSDCF und RMSDSN-Werte zeigen in Abbildung 54 erstaunlicherweise leicht bessere Werte für kleinere  $\Delta_{\max}(1)$ , obwohl der Effekt fast vollständig im Bereich der Fehlergrenzen liegt. Unter 5 (50 %) zeigen RMSG und RMSDCF eine klare Qualitätsverschlechterung, welche durch visuelle Inspektion auf eine große Zahl degenerierter Netzkarten zurückzuführen ist. Der RMSDSN-Wert zeigt auch für diese degenerierten Netzkarten gute Werte an und ist zur Erkennung derartiger Netzfehler offensichtlich nicht geeignet. Da die Anzahl der zu trainierenden Strukturen und damit die Rechenzeit quadratisch von der Größe der Trainingsumgebung abhängen, lohnt sich eine Optimierung

von  $\Delta_{\max}(1)$ . Der in Vorversuchen ermittelte Kompromiss von 80% ist als Standardwert in das Programm NN.EXE (5.3.15) eingetragen.

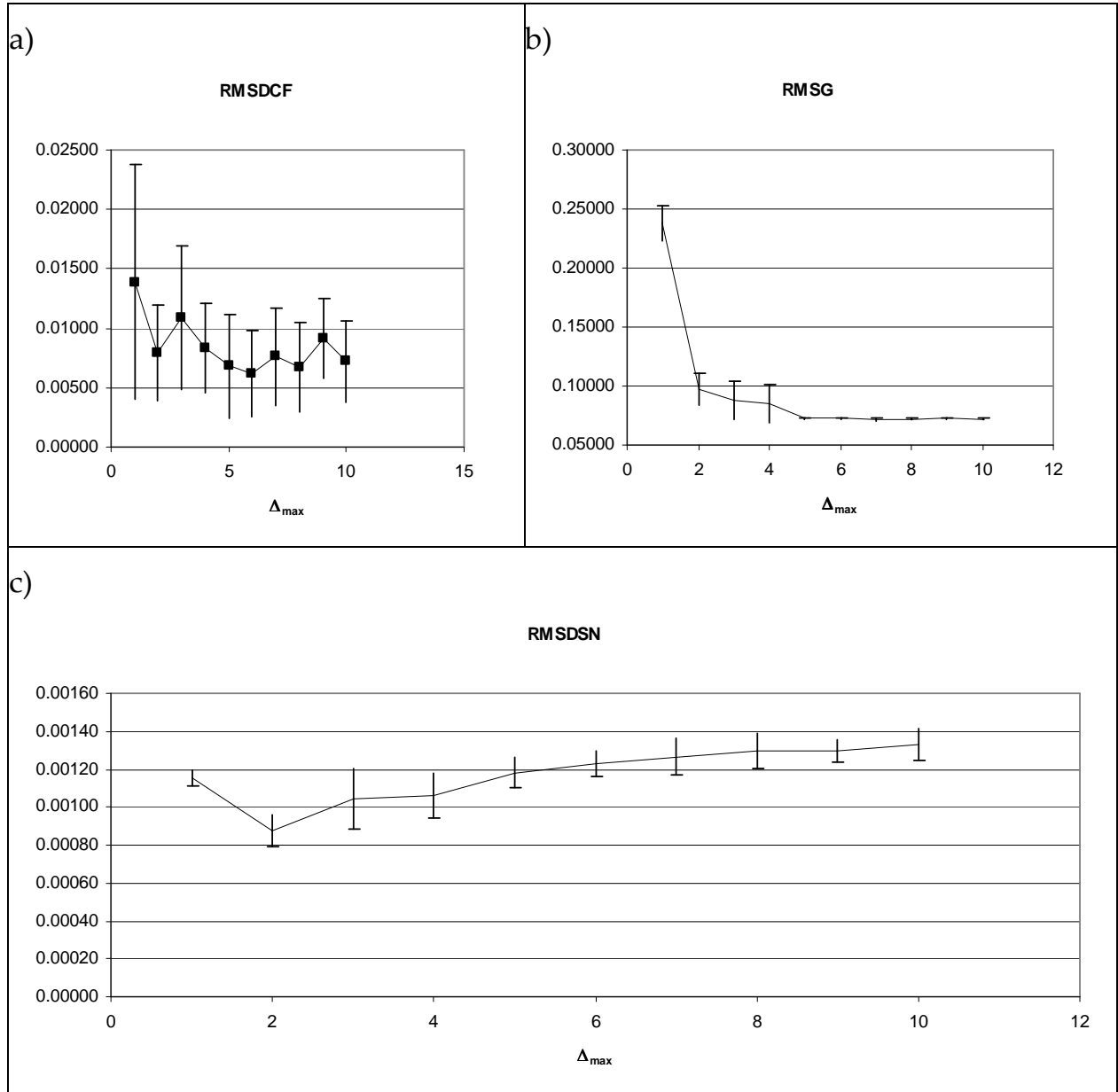


Abbildung 54: Variation von  $\Delta_{\max}(1)$ , a) RMSDCF, b) RMSG, c) RMSDSN

**2.8.1.5.4 Variation von  $\eta(\xi)$**

Tabelle 18: Variation der Zeitabhängigkeit des Adaptionfaktors,  $\eta(\xi)$  nach Gleichung (2.17)

$\Delta_{\max}(1)$	$\bar{\sigma}$ RMSDCF	$\sigma$	relerr%	min	max	ndata
a) linear	0.00749	0.00331	44.2	0.00335	0.01213	10
b) quadratisch	0.00766	0.00371	48.4	0.00323	0.01447	10
$\Delta_{\max}(1)$	$\bar{\sigma}$ RMSG	$\sigma$	relerr%	min	max	ndata
a) linear	0.07230	0.00076	1.1	0.07105	0.07352	10
b) quadratisch	0.07186	0.00085	1.2	0.07061	0.07301	10

$\Delta_{\max}(1)$	$\emptyset$ RMSDSN	$\sigma$	relerr%	min	max	ndata
a) linear	0.00134	0.00009	6.5	0.00124	0.00154	10
b) quadratisch	0.00227	0.00030	13.2	0.00180	0.00271	10

Die grafische Auswertung in Abbildung 55 ergibt im Rahmen der Fehlergrenzen keine qualitativen Unterschiede für eine der beiden Varianten. Der RMSDSN-Wert zeigt leichte Vorteile für die in dieser Arbeit als Standardvariante verwendete Form a) in Gleichung (2.17). Die anderen Parameter zeigen im Rahmen der Fehlergrenzen keine signifikanten Unterschiede.

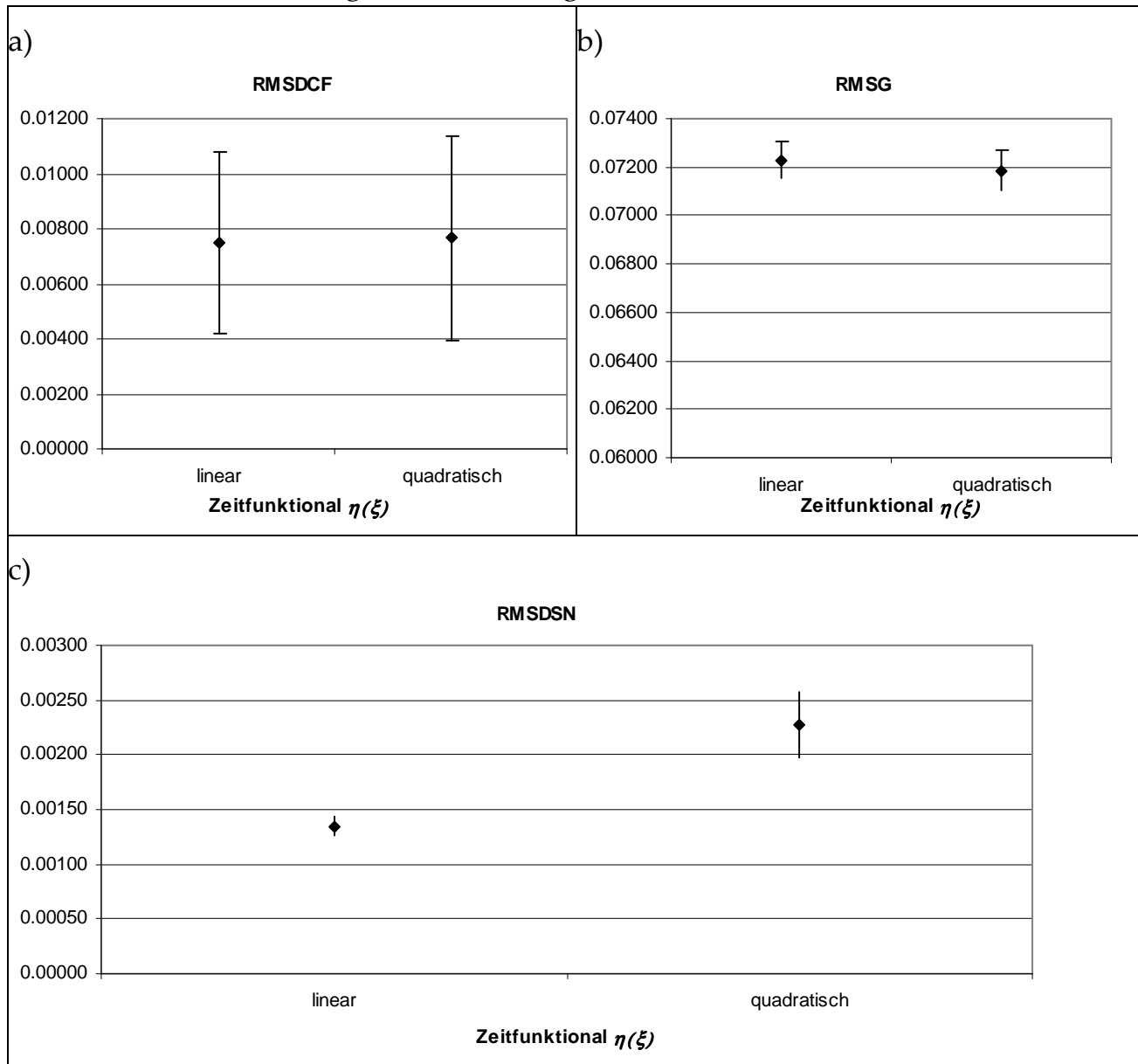


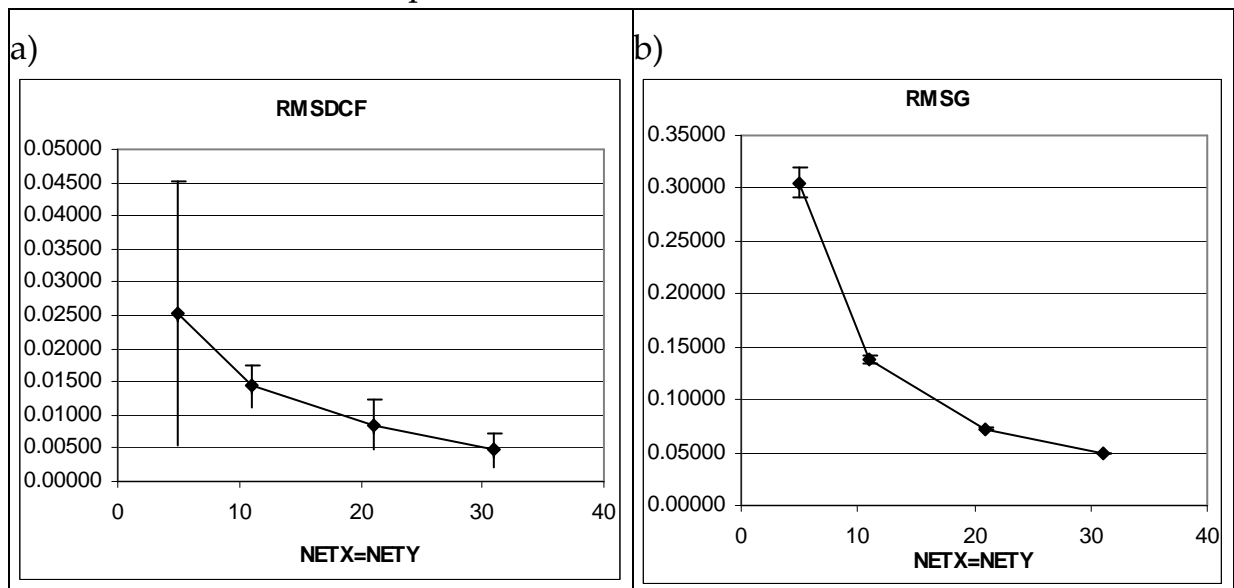
Abbildung 55: Variation von  $\eta(\xi)$  nach Gleichung (2.17)

## 2.8.1.5.5 Variation von NETX, NETY

Tabelle 19: Variation der Netzgröße NETX=NETY

NETX=NETY	Ø RMSDCF	$\sigma$	relerr%	min	max	ndata
31	0.00472	0.00257	54.48	0.00033	0.00927	10
21	0.00854	0.00368	43.11	0.00298	0.01314	10
11	0.01432	0.00319	22.29	0.01111	0.02030	10
5	0.02529	0.01996	78.93	0.01563	0.07875	10
NETX=NETY	Ø RMSG	$\sigma$	relerr%	min	max	ndata
31	0.04884	0.00034	0.70	0.04821	0.04919	10
21	0.07267	0.00057	0.79	0.07207	0.07366	10
11	0.13793	0.00308	2.23	0.13309	0.14208	10
5	0.30538	0.01497	4.90	0.29554	0.33511	10
NETX=NETY	Ø RMSDSN	$\sigma$	relerr%	min	max	ndata
31	0.00066	0.00005	7.15	0.00059	0.00074	10
21	0.00134	0.00007	5.55	0.00123	0.00146	10
11	0.00362	0.00026	7.19	0.00328	0.00409	10
5	0.01282	0.00051	4.00	0.01229	0.01402	10

Die grafische Abbildung zeigt, dass generell größere Netzkarten bessere Deskriptoren ergeben. Die Kurven scheinen einer negativen Exponentialfunktion  $c1 \cdot \exp(-c2)$  zu folgen, so dass sehr schnell die Qualitätssteigerung in keinem Verhältnis zur quadratisch wachsenden Rechenzeitsteigerung mehr steht. Eine minimale Größe der Netzkarte scheint zu existieren, wenn man davon ausgeht, dass die Streuung der RMSDCF-Werte deutlich kleiner als der absolute RMSDCF-Wert sein sollte. Im vorliegenden Fall liegt die ermittelte Grenze damit bei  $\text{NETX}=\text{NETY} \geq 10$ , was einem Verhältnis Neuronen : Konformationen = 100 : 6 entspricht.



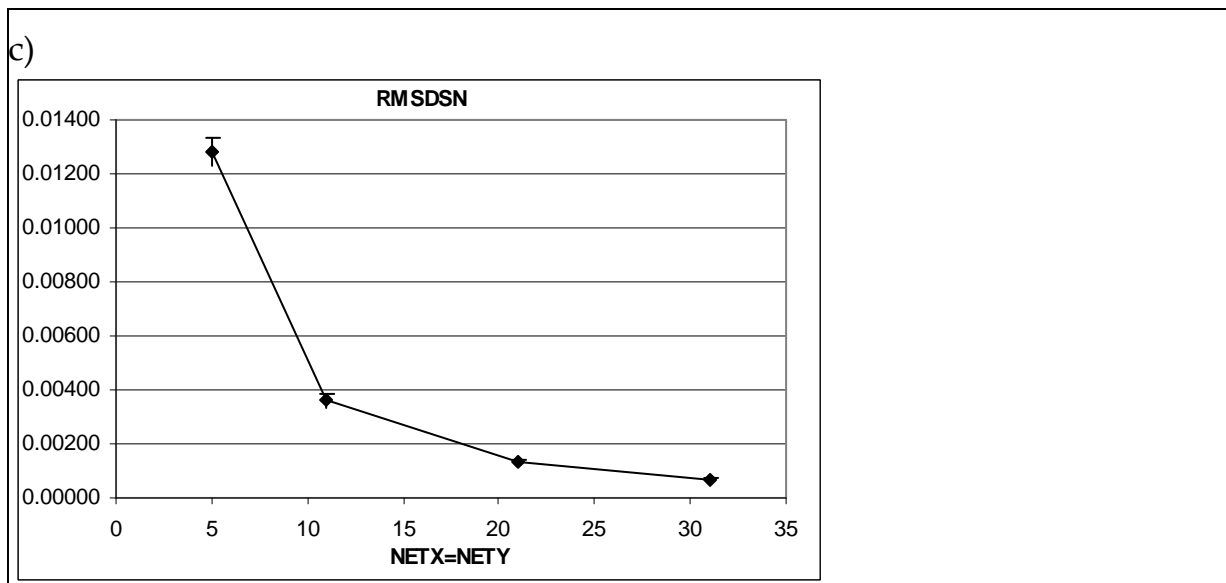


Abbildung 56: Variation der Netzgröße NETX=NETY

### 2.8.1.5.6 Variation der Abstandsfunktion $\vartheta(\zeta)$

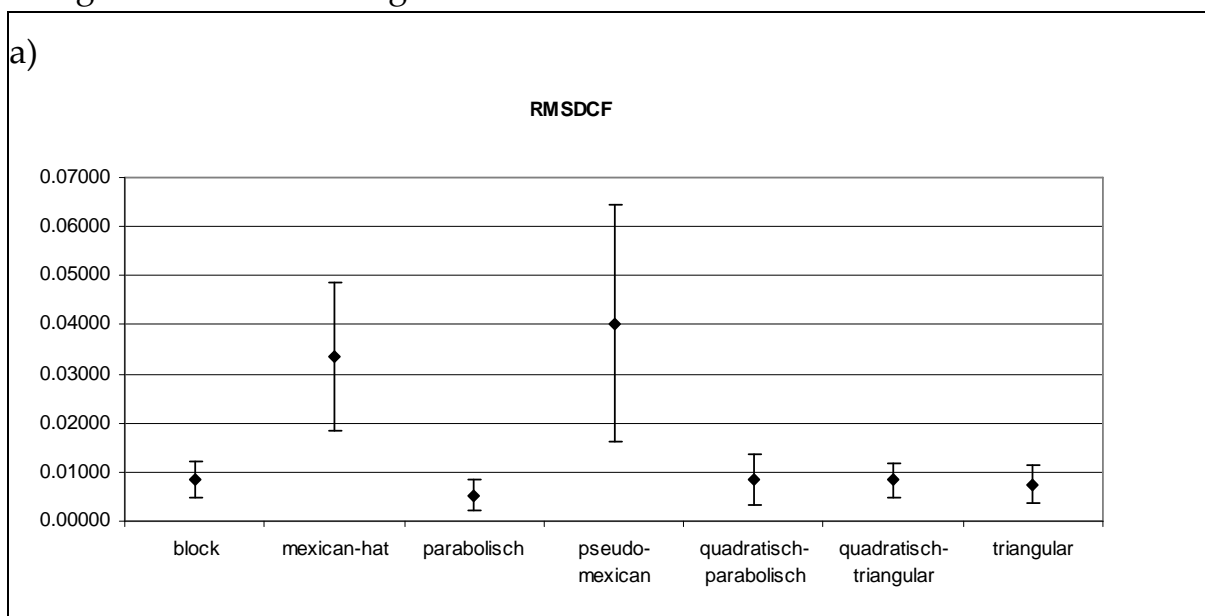
Tabelle 20: Variation der Abstandsfunktion  $\vartheta(\zeta)$  nach Gleichung (2.18)

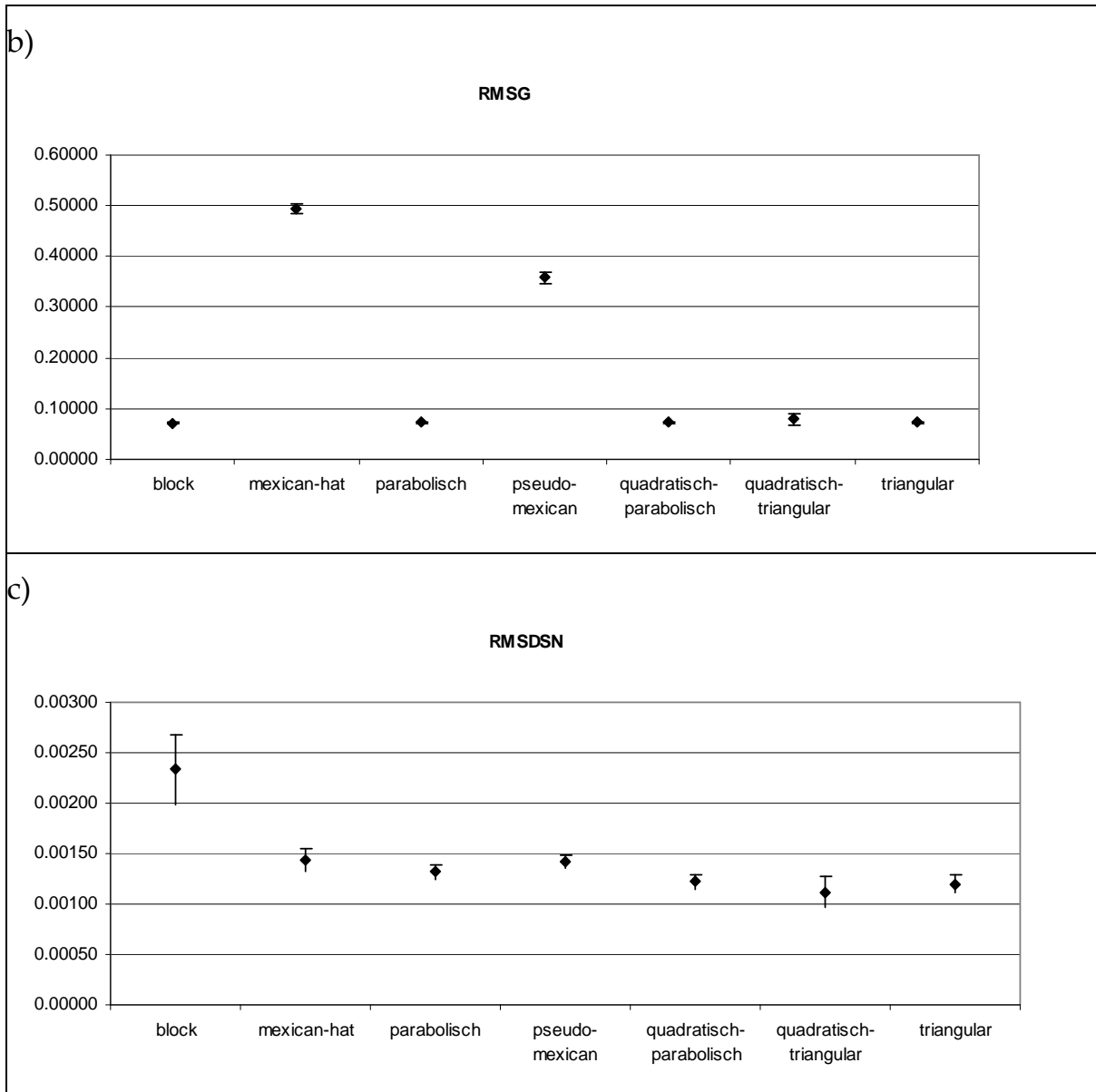
$\vartheta(\zeta)$	$\emptyset$ RMSDCF	$\sigma$	re- lerr%	min	max	nda- ta
<i>block</i>	0.00855	0.00361	42.27	0.00518	0.01406	10
<i>Mexican hat</i>	0.03345	0.01518	45.39	0.01540	0.06092	10
parabolisch	0.00530	0.00299	56.32	0.00298	0.01101	10
<i>pseudo Mexican</i>	0.04030	0.02406	59.71	0.00226	0.07272	10
quadratisch- parabolisch	0.00848	0.00500	58.96	0.00323	0.01934	10
quadratisch- triangular	0.00833	0.00350	42.02	0.00234	0.01383	10
triangular	0.00755	0.00382	50.55	0.00264	0.01329	10
$\vartheta(\zeta)$	$\emptyset$ RMSG	$\sigma$	re- lerr%	min	max	nda- ta
<i>block</i>	0.07181	0.00045	0.62	0.07108	0.07255	10
<i>Mexican hat</i>	0.49345	0.01039	2.10	0.46990	0.50523	10
parabolisch	0.07238	0.00085	1.18	0.07124	0.07370	10
<i>pseudo Mexican</i>	0.35813	0.01160	3.24	0.34309	0.37700	10
quadratisch- parabolisch	0.07236	0.00043	0.59	0.07160	0.07296	10
quadratisch- triangular	0.07895	0.01203	15.24	0.07269	0.10220	10
triangular	0.07256	0.00053	0.73	0.07147	0.07339	10



$\mathfrak{S}(\zeta)$	$\emptyset$ RMSDSN	$\sigma$	re- lerr%	min	max	nda- ta
<i>block</i>	0.00234	0.00035	14.84	0.00190	0.00284	10
<i>Mexican hat</i>	0.00144	0.00011	7.59	0.00126	0.00166	10
parabolisch	0.00132	0.00008	5.74	0.00124	0.00148	10
<i>pseudo Mexican</i>	0.00142	0.00007	4.78	0.00134	0.00157	10
quadratisch- parabolisch	0.00122	0.00008	6.44	0.00113	0.00135	10
quadratisch- triangular	0.00112	0.00015	13.72	0.00083	0.00126	10
triangular	0.00120	0.00009	7.31	0.00107	0.00131	10

Die grafische Auswertung in Abbildung 57 ergibt als eindeutiges Ergebnis, dass „*Mexican hat*“-artige Funktionen (siehe Abbildung 24), wie sie ursprünglich von Kohonen propagiert wurden, sich nicht für das Training von Konformerensembles eignen. Die RMSDCF- und RMSG-Werte sowie deren Streuung sind zu hoch. Dies korreliert mit einer hohen Anzahl degenerierter Netzkarten, wie durch visuelle Inspektion festgestellt werden kann. Die Qualitätsunterschiede zwischen *block*, parabolischer, quadratisch-parabolischer, quadratisch-triangularer und triangularer Funktion sind weitgehend vernachlässigbar.





**Abbildung 57: Variation der Abstandsfunktion  $\vartheta(\zeta)$**

### 2.8.1.5.7 Fazit

Die verwendeten Standardnetzparameter scheinen bereits eine gute Basis darzustellen. Grundsätzlich sollten immer mehrere Netzkarten mit den gleichen Netzparametern trainiert werden, um reproduzierbare Daten zu erhalten und degenerierte Ausreißer erkennen zu können. Für die einzelnen Netzparameter gibt es folgende Erkenntnisse, welche eventuell spezifisch für Butan sind:

- *epochs*: ermittelte minimale Trainingsschritte nach Kapitel 2.5.4  
 $[f_0=0.15, \Delta_{\max}(1)=10(100\%), \eta(\xi)=\text{linear}, \text{NETX}/\text{NETY}=21,$   
 $\vartheta(\zeta)=\text{parabolisch}]$   
 MTS(RMSG)=180000  
 MTS(RMDSN)=720000

- $f_0$ :  
Hier scheint sich der sehr konservative Vorgabewert von 0.15 eher an der unteren Grenze des möglichen Wertebereichs zu befinden. Zumindest in obiger Analyse wäre ein optimaler Wert eher um die 0.8 zu finden
- $\Delta_{\max}(1)$   
Entgegen den Erwartungen scheint nicht nur das Rechenzeitoptimum, sondern auch das Qualitätsoptimum nicht bei 100% wie in der Voreinstellung, sondern bei Werten von  $\geq 50\%$  des Maximalwertes zu liegen. Ob dies bei komplexeren Hyperflächen so bleibt, ist fraglich. Auch eine Kreuzkorrelation mit der Abstandsfunktion ist zu vermuten, wird hier aber nicht untersucht.
- $\eta(\xi)$   
Kein klarer Sieger; die lineare Funktion erweist sich als marginal überlegen („quadratic learning decay=off“)
- NETX/NETY  
Die optimale Netzgröße ist abhängig von der darzustellenden Hyperfläche. Der Zusammenbruch der Netzprojektion erfolgt in obiger Analyse bei 25 Neuronen, was eine Untergrenze von ca. 9 Neuronen pro darzustellendes Molekülminimum bedeutet. In der Praxis kann die Netzkarte so gewählt werden, wie es für die weitere Auswertung am besten handhabbar ist, da eine nicht verlustfreie Reduktion der Redundanz durchaus ein Feature darstellt.
- $\vartheta(\zeta)$   
Als Abstandsfunktion kommen in vorliegender Anwendung der Kohonnennetze keinesfalls Varianten der „Mexican hat“-Funktion in Betracht. Gleichbleibend gute Ergebnisse werden mit allen „dreiecksähnlichen“ Funktionen erzielt, also parabolisch, quadratisch-parabolisch, quadratisch-triangular und triangular. Die *block* Funktion kann eher den „dreiecksähnlichen“ Funktionen zugeordnet werden.

#### **2.8.1.6 Beurteilung der Qualitätskriterien**

Aus obigen Messreihen in Kapitel 2.8.1.5 in Verbindung mit einer visuellen Inspektion der erhaltenen Netzkarten lässt sich eine Bewertung der verwendeten Qualitätskriterien ableiten.

## RMSDCF:

Der RMSDCF-Wert ist am aufwendigsten zu ermitteln und erfordert eine Clusteranalyse der Trajektorie sowie der Netzkarte und ist damit unter Umständen nicht deterministisch. Von allen drei Kriterien ist er dasjenige mit dem größten relativen Fehler. Dies ist ein direkter Hinweis, dass die Flächenverteilung der Netzkarten nur qualitativ, nicht aber quantitativ, als Analyse-methode für die Anteile von einzelnen Konformationen an der Gesamtheit aller Eingabedaten erhalten kann. Degenerierte Netzkarten können durch einen signifikant erhöhten RMSDCF-Wert meist erkannt werden.

## RMSG:

Der RMSG-Wert zeigt von allen Qualitätsparametern den kleinsten relativen Fehler und ist am billigsten zu berechnen. Des Weiteren ist ein hoher RMSG-Wert die sicherste Methode, degenerierte Netzkarten im Vergleich zu nicht degenerierten zu erkennen. Folgerichtig wird der RMSG-Wert im Verlaufe des Netztrainings stichprobenweise berechnet und in der .net.log Datei protokolliert. Dies ermöglicht es, schon während des Netztrainings die Qualität der erhaltenen Karten zu beurteilen. Eine typische Kurve des Trainings einer 10x10 Netzkarte ist in Abbildung 58 dargestellt. Erkennbar ist eine stufenweise Abnahme der Varianz, welche hauptsächlich gekoppelt ist mit der Verringerung der Größe der trainierten Nachbarschaft.

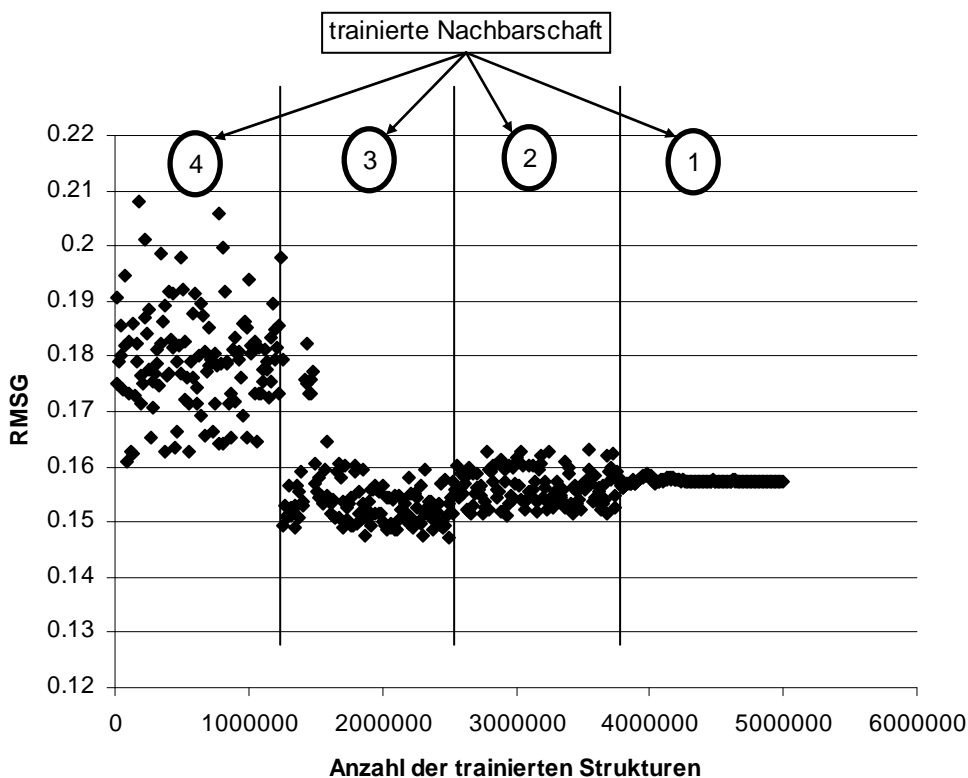


Abbildung 58: typischer Verlauf der RMSG-Werte während des Trainings

Ebenso nehmen die RMSG-Werte nach anfänglicher Abnahme in Stufen wieder zu. Obige Abbildung ist in voller Übereinstimmung mit Abbildung 51. Es ist an dieser Stelle naheliegend, den Verlauf der RMSG-Werte als Kriterium für die Verringerung der Trainingsdistanz und letztendlich als Abbruchkriterium zu verwenden. Eine probeweise Implementierung („*fixed nr of epochs*“ unselektiert in Abbildung 37b)) ergibt keine guten Ergebnisse und scheitert zum einen an den Schwierigkeiten, aus dem Trend der RMSG-Werte, das Erreichen einer Stufe zu erkennen. Zum anderen ist es grundsätzlich fragwürdig, ob der RMSG-Wert ein gutes Abbruchkriterium darstellt, da er im Verlaufe des Trainings auch wieder ansteigt, bzw. konstant bleibt bei weiter fallendem RMSDSN-Wert (siehe Abbildung 51).

RMSDSN:

Der relative Fehler der RMSDSN-Werte ist relativ klein, das Kriterium damit gut reproduzierbar. Die Berechnung des RMSDSN-Wertes liegt vom Aufwand her zwischen den anderen beiden Kriterien. Im Gegensatz zum RMSDCF-Wert muss keine zusätzliche Cluster-Analyse durchgeführt werden. Zur Ermittlung müssen jedoch die Abstände aller Trainingsdaten zu allen Netzkarten errechnet werden, was in der Regel eine um Größenordnungen längere Rechenzeit bedeutet als bei der Bestimmung des RMSG-Wertes.

Der RMSDSN-Wert scheint grundsätzlich nicht geeignet zu sein, degenerierte Netzkarten zu erkennen. Dies kann anhand von Abbildung 57 nachgewiesen werden. Andererseits entspricht ein niedriger RMSDSN-Wert einer guten Repräsentation der trainierten Werte durch die Netzkarte, was einem der geforderten Qualitätskriterien exakt entspricht. In Abbildung 51 ist zu erkennen, dass dieses Qualitätskriterium durchaus unabhängig von anderen Kriterien variieren kann.

Zusammenfassend lassen sich die Qualitätskriterien in zwei Gruppen aufteilen. RMSDCF und RMSG sind nützlich zur Untersuchung der erhaltenen Netzkarten auf Degeneriertheit. Der RMSDCF-Wert scheint damit prinzipiell überflüssig zu sein. Bei vergleichbarem RMSG-Wert als primärem Qualitätskriterium kann es trotzdem zu variierenden RMSDSN-Werten kommen, so dass der RMSDSN-Wert als sekundäres Qualitätskriterium herangezogen werden sollte.

### **2.8.1.7 MD Butan, Training der Diederwinkel**

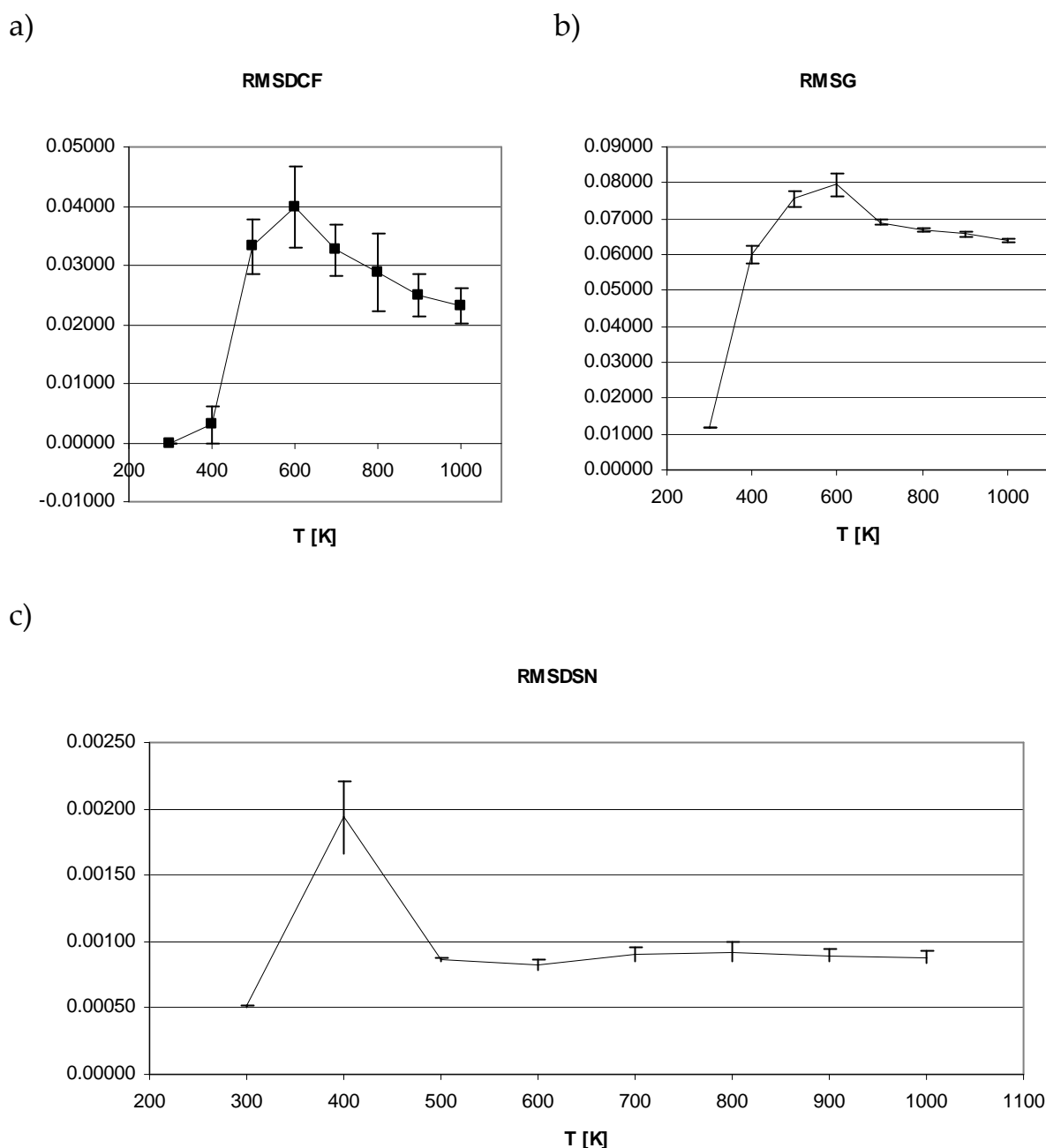
Um die Stabilität des Netztrainings auch für nicht optimale Eingabedaten zu überprüfen, wird im Folgenden eine Gruppe Moleküldynamikrechnungen

trainiert, welche bei unterschiedlichen Temperaturen simuliert wurden (Kraftfeld HYPERCHEM: MM+, in vacuo, 20000 Strukturen, Simulationsdauer 10000 ps, Aufruf von FIX\_ROTATION.EXE bei je 1 ps, *step size* 0.001 ps, *datafreq*=1/500 *data steps*<sup>-1</sup>). Es werden also 10000 ps mit einer Schrittweite von 1fs simuliert, jede fünfhundertste Struktur wird im Konformerensemble für das Netztraining gespeichert. Die erhaltenen Daten sind in Tabelle 21 angeführt. Die Netzparameter betragen: [25x25 Neuronen; 100 *epochs*;  $\Delta_{\max}(1)=12$ ;  $f_0=0.5$ ; Abstandsfunktion parabolisch, Training des zentralen Diederwinkels].

**Tabelle 21: Butan MD-Simulationen bei unterschiedlichen Temperaturen, Training des zentralen Diederwinkels**

T [K]	Ø RMSDCF	$\sigma$	relerr%	min	max	ndata
1000	0.02317	0.00305	13.18	0.01952	0.02981	10
900	0.02488	0.00361	14.50	0.01942	0.03118	10
800	0.02882	0.00655	22.71	0.02145	0.04179	10
700	0.03256	0.00426	13.09	0.02557	0.03919	10
600	0.03991	0.00687	17.21	0.03200	0.05475	10
500	0.03324	0.00466	14.01	0.02760	0.04120	10
400	0.00303	0.00320	105.48	0.00069	0.01176	10
300	0.00000	0.00000	0.00	0.00000	0.00000	10
T [K]	Ø RMSG	$\sigma$	relerr%	min	max	ndata
1000	0.06398	0.00054	0.84	0.06333	0.06491	10
900	0.06566	0.00057	0.87	0.06497	0.06682	10
800	0.06685	0.00057	0.85	0.06605	0.06775	10
700	0.06902	0.00069	1.00	0.06803	0.06997	10
600	0.07952	0.00331	4.17	0.07507	0.08391	10
500	0.07562	0.00226	2.99	0.07324	0.07889	10
400	0.06000	0.00242	4.03	0.05843	0.06676	10
300	0.01195	0.00010	0.81	0.01186	0.01219	10
T [K]	Ø RMSDSN	$\sigma$	relerr%	min	max	ndata
1000	0.00088	0.00004	4.90	0.00083	0.00096	10
900	0.00090	0.00004	4.95	0.00082	0.00095	10
800	0.00092	0.00007	7.63	0.00079	0.00103	10
700	0.00090	0.00005	5.72	0.00080	0.00097	10
600	0.00082	0.00004	4.86	0.00077	0.00090	10
500	0.00086	0.00002	2.02	0.00083	0.00088	10
400	0.00194	0.00027	13.86	0.00119	0.00210	10
300	0.00051	0.00000	0.86	0.00050	0.00052	10

Die absoluten Werte der Qualitätskriterien liegen im gleichen Größenbereich wie bei den Netzkarten aus der synthetischen Hyperfläche. RMSDCF=0.008, RMSG=0.07 RMSDSN=0.001 (siehe Kapitel 2.8.1.5).



**Abbildung 59: Variation der Temperatur – Butan MD, Training des zentralen Diederwinkels**

Die grafische Auftragung der Qualitätsparameter in Abbildung 59 zeigt einen eigentümlichen Verlauf mit einem Maximum der RMSDCF-/RMSG-Werte bei ca. 600 K, bzw. einem Maximum der RMSDSN-Werte bei ca. 400 K. Dies lässt sich erklären durch ein Verschieben der Gleichgewichtslage der Molekülstrukturen hin zu höherer Energie bei höherer Temperatur nach Boltzmannstatistik. Während bei 300 K das Molekül im Rahmen der *in-vacuo*-Simulation in der *anti*-Konformation gefangen ist, kommt es ab 400 K zu Anteilen aller drei Konformationen, *anti*, *gauche<sup>+</sup>*, *gauche<sup>-</sup>*. Dies kann je nach Anteilen der einzelnen Konformationen zu den in Abbildung 30 dargestellten Netztopologien führen, welche anhand ihrer RMSG-Werte gut unterschieden werden können.

Wie aus Abbildung 59b), der grafischen Auftragung des RMSG-Wertes, am besten zu erkennen ist, werden ab 700 K reproduzierbar Netztopologien einer einzigen Art erhalten, welche visuell als identisch zu Abbildung 30a) identifiziert werden.

Es ist zu beachten, dass die Simulationen bei höheren Temperaturen anscheinend besser äquilibrierte Eingabedaten und damit bessere Netzkarten liefern. Dies steht im Gegensatz zu anderen Auswertemethoden, bei denen „thermisches Rauschen“ oftmals eher zu schlechterer Qualität führt. Beispiele für die erhaltenen Netzkarten sind in Abbildung 60 dargestellt. Die Kartentopologien sind bereits aus Abbildung 43 und Abbildung 53 bekannt und sind dort ausführlich analysiert.



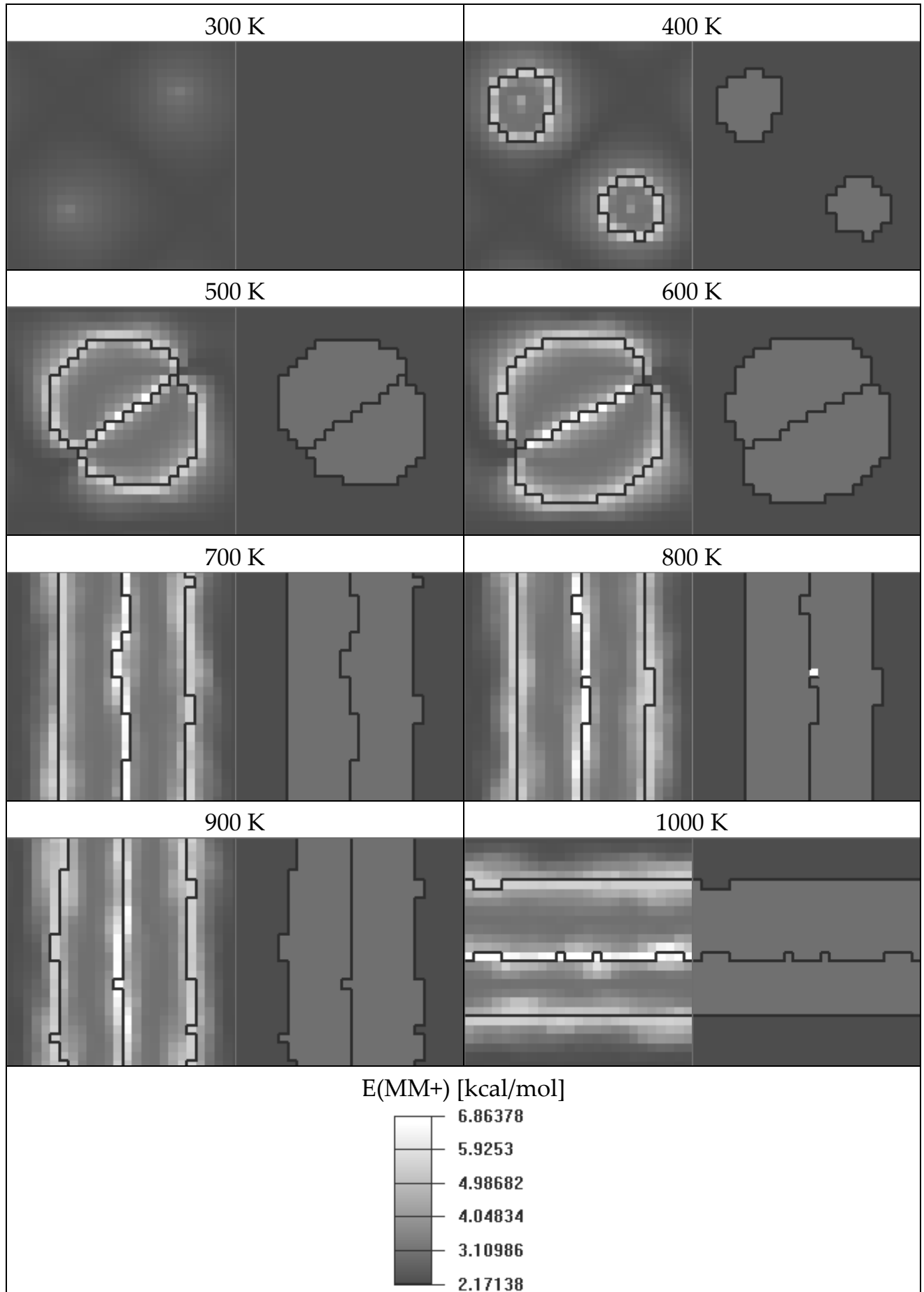


Abbildung 60: Beispielnetz Karten Gridhyperfläche – Variation der Temperatur – Butan MD, Training des zentralen Diederwinkels, links partiell optimiert nach COLORIX1, rechts vollständig optimiert nach COLORIX2.

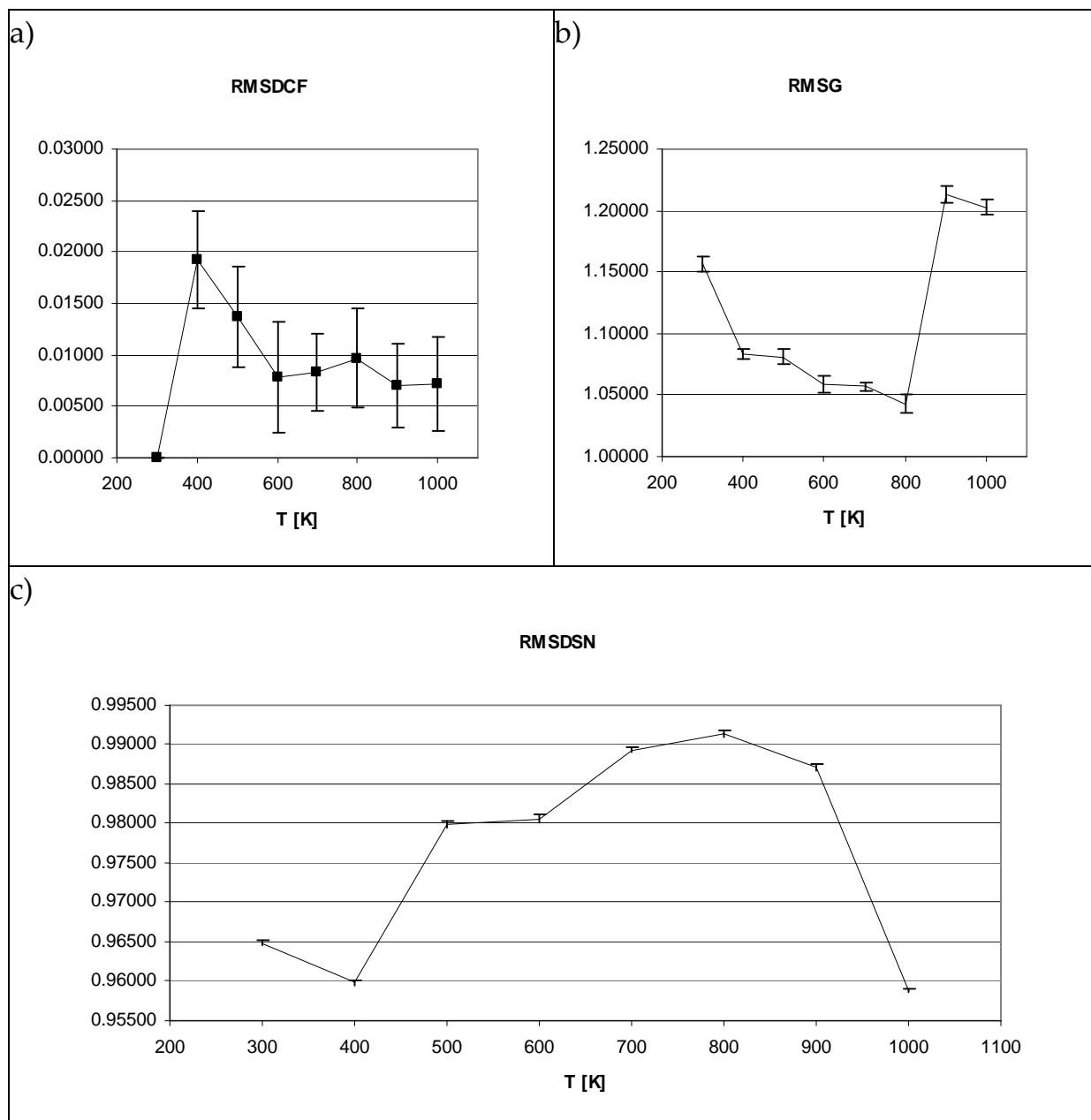
### 2.8.1.8 MD Butan, Training aller Distanzen

Das Training aller Distanzen, inklusive der C-C, C-H und H-H Abstände, ergibt eine komplexere Hyperfläche, da die Drehung der endständigen Methylgruppen über die Distanzmatrix unterscheidbar wird. Dies führt zu Numerationsisomeren (Konformere, die ausschließlich aufgrund der Nummerierung der Atome unterscheidbar sind). Die statistischen Daten des Netztrainings der Moleküldynamiken aus Kapitel 2.8.1.7 sind in Tabelle 22 aufgeführt. [25x25 Neuronen; 100 *epochs*;  $\Delta_{\max}(1)=12$ ;  $f_0=0.5$ ; Abstandsfunktion parabolisch, Training aller Abstände]

**Tabelle 22: Butan MD-Simulationen bei unterschiedlichen Temperaturen, Training der kompletten Distanzmatrix**

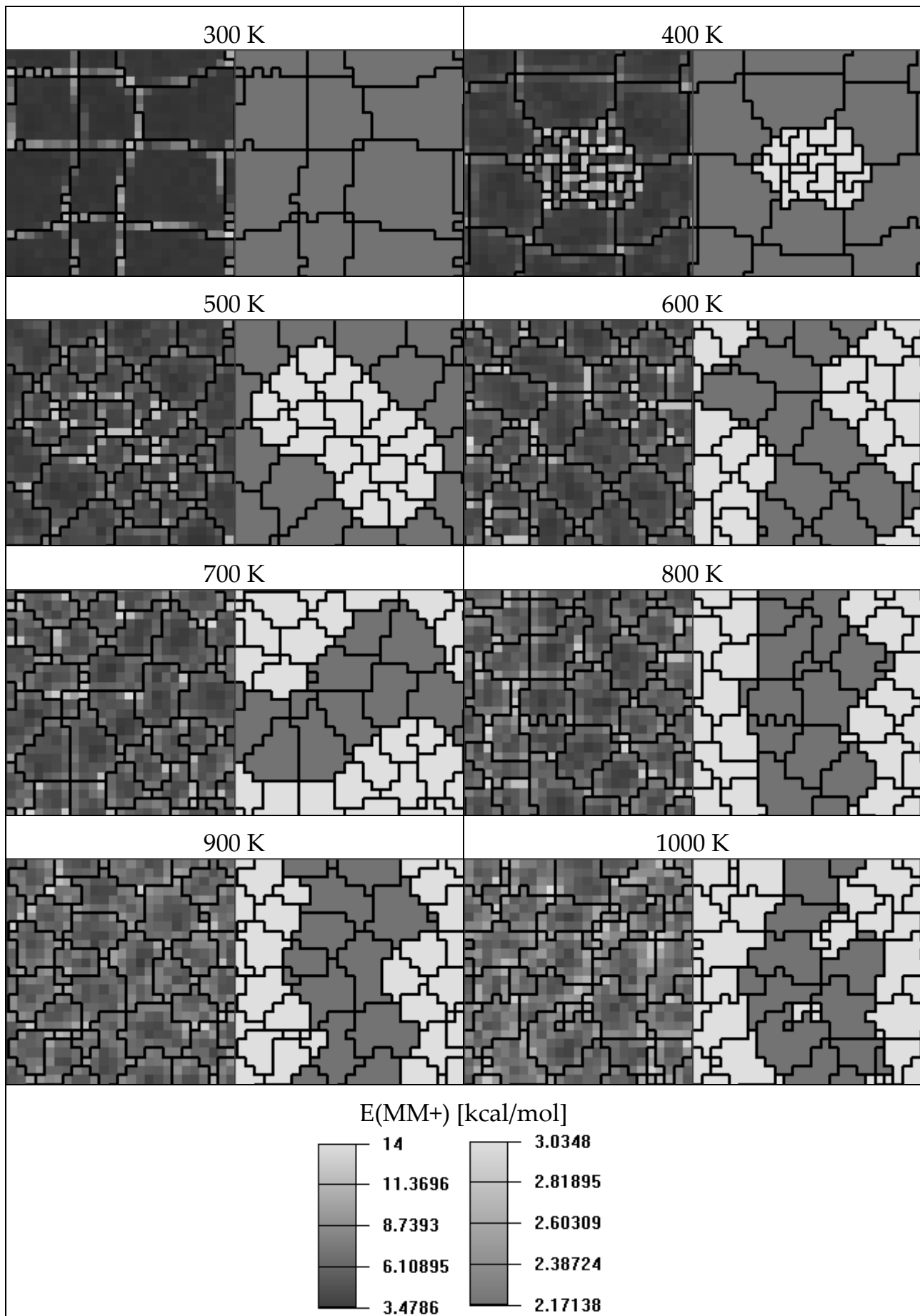
T [K]	Ø RMSDCF	$\sigma$	relerr%	min	max	ndata
1000	0.00718	0.00451	62.77	0.00322	0.01665	10
900	0.00695	0.00406	58.44	0.00151	0.01397	10
800	0.00966	0.00478	49.49	0.00237	0.01773	10
700	0.00826	0.00376	45.50	0.00348	0.01530	10
600	0.00781	0.00544	69.66	0.00194	0.01956	10
500	0.01370	0.00487	35.52	0.00778	0.02346	10
400	0.01924	0.00473	24.59	0.01224	0.02688	10
300	0.00000	0.00000	0.00	0.00000	0.00000	10
T [K]	Ø RMSG	$\sigma$	relerr%	min	max	ndata
1000	1.20275	0.00605	0.50	1.19365	1.21110	10
900	1.21348	0.00705	0.58	1.20398	1.22503	10
800	1.04278	0.00735	0.70	1.03252	1.05405	10
700	1.05704	0.00340	0.32	1.05302	1.06337	10
600	1.05886	0.00672	0.63	1.04799	1.07108	10
500	1.08110	0.00573	0.53	1.07175	1.08943	10
400	1.08281	0.00417	0.38	1.07714	1.08912	10
300	1.15651	0.00649	0.56	1.14848	1.16987	10
T [K]	Ø RMSDSN	$\sigma$	relerr%	min	max	ndata
1000	0.95883	0.00029	0.03	0.95826	0.95923	10
900	0.98710	0.00043	0.04	0.98651	0.98771	10
800	0.99144	0.00046	0.05	0.99069	0.99225	10
700	0.98935	0.00038	0.04	0.98902	0.99006	10
600	0.98062	0.00050	0.05	0.97996	0.98160	10
500	0.97987	0.00050	0.05	0.97901	0.98054	10
400	0.95985	0.00036	0.04	0.95930	0.96042	10
300	0.96488	0.00039	0.04	0.96403	0.96536	10

Die insgesamt höheren absoluten RMSG- und RMSDCF-Werte sind auf den größeren Satz an abzubildenden Konformationen zurückzuführen. In Abbildung 61 sind die Qualitätsparameter des Netztrainings grafisch aufgetragen.



**Abbildung 61: Variation der Temperatur – Butan MD, Training der kompletten Distanzmatrix**

Abbildung 62 zeigt deutlich den unterschiedlichen Verlauf der MD-Simulationen bei verschiedenen Temperaturen. Während bei 300 K aufgrund der Rotation der endständigen Methylgruppen ausschließlich 9 Numerationskonformere vorkommen, zeigen die Netzkarten ab 400 K einen steigenden Anteil an *gauche*-Konformationen. Die Korrelation von Abbildung 61 mit Abbildung 62 zeigt, dass die Topologie der Netzkarte nur einer der Faktoren ist, welche in die Qualitätsparameter eingehen. Während in unten stehenden Netzkarten ab 600 K keine wesentlichen topologischen Änderungen zu sehen sind, kommt es laut Abbildung 61c bei 900 K zu einer rapiden Verschlechterung des RMSG-Wertes. Dieser kann wahrscheinlich durch das thermische Rauschen erklärt werden, zumindest deutet die subjektiv unruhigere COLORIX1 Netzkarte darauf hin.



**Abbildung 62:** Beispielnetzarten Gridhyperfläche – Variation der Temperatur – Butan MD, Training der kompletten Distanzmatrix, links partiell optimiert nach COLORIX1, rechts vollständig optimiert nach COLORIX2

Die Netzkarten für das Training von kompletten Distanzmatrizen mit den resultierenden Numerationssisomeren sind in der Regel zu detailreich, um von Nutzen zu sein. So führt der weiter oben erwähnte Effekt der Numerationssisomeren z.B. dazu, dass die helle *gauche*-Fläche in Abbildung 62 in 18 Unterbereiche geclustert wird (2 *gauche* Konformationen, je drei Stellungen der endständigen Methylgruppen).

Distanzmatrix-Netzarten können jedoch die Interpretation von Koaleszenzeffekten in NMR-Spektren erleichtern, da diese der Aufspaltung einer Netzhyperfläche durch Verwendung von Distanzmatrizen entsprechen.

### 2.8.1.9 MD Butan, Training der Schweratomabstandsmatrix

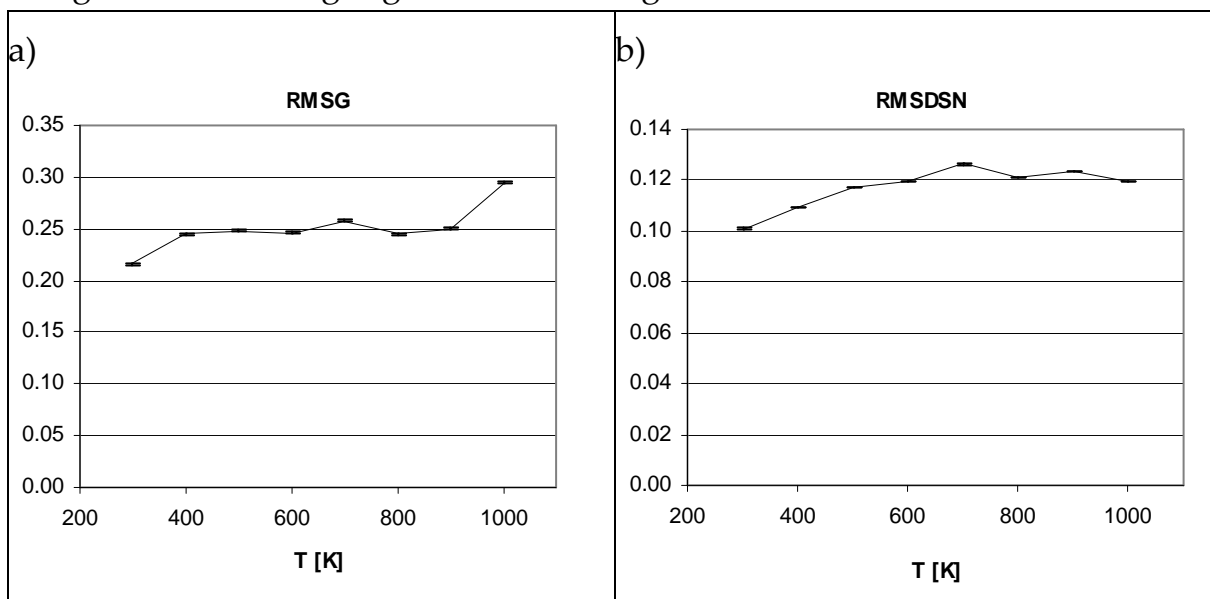
Das Training der Schweratomabstandsmatrix, in diesem Falle also aller C-C Abstände, ergibt eine vereinfachte Hyperfläche, da die beiden *gauche*-Konformationen auf Grund von C-C Abständen nicht unterscheidbar sind. Allgemein können bei den linearen Alkanen „Dieder-Enantiomere“ nicht aufgrund der Schweratomabstandsmatrix auseinander gehalten werden. Entsprechend kann der RMSDCF-Wert nicht korrekt berechnet werden, da die verwendeten Methoden sehr wohl „Dieder-Enantiomere“ differenzieren. Der RMSDCF-Wert wird für die linearen Alkane folglich nicht angegeben.

Die statistischen Daten des Netztrainings der Moleküldynamiken aus Kapitel 2.8.1.7 sind in Tabelle 23 aufgeführt. [25x25 Neuronen; 20 *epochs*;  $\Delta_{\max}(1)=12$ ;  $f_0=0.5$ ; Abstandsfunktion parabolisch, Training der C-C Abstände]

**Tabelle 23: Butan MD-Simulationen bei unterschiedlichen Temperaturen, Training der Schweratommatrix**

T [K]	Ø RMSG	$\sigma$	relerr%	min	max	ndata
1000	0.29425	0.00075	0.25	0.29244	0.29511	10
900	0.24962	0.00111	0.45	0.24752	0.25177	10
800	0.24540	0.00099	0.40	0.24314	0.24659	10
700	0.25783	0.00087	0.34	0.25648	0.25924	10
600	0.24631	0.00102	0.41	0.24505	0.24802	10
500	0.24848	0.00074	0.30	0.24723	0.24993	10
400	0.24509	0.00099	0.41	0.24407	0.24725	10
300	0.21596	0.00095	0.44	0.21448	0.21778	10
T [K]	Ø RMSDSN	$\sigma$	relerr%	min	max	ndata
1000	0.11944	0.00009	0.08	0.11934	0.11961	10
900	0.12335	0.00018	0.15	0.12312	0.12363	10
800	0.12132	0.00011	0.09	0.12117	0.12154	10
700	0.12629	0.00015	0.12	0.12613	0.12654	10
600	0.11938	0.00013	0.11	0.11919	0.11957	10
500	0.11734	0.00012	0.10	0.11713	0.11757	10
400	0.10943	0.00018	0.17	0.10920	0.10978	10
300	0.10104	0.00016	0.16	0.10076	0.10131	10

Die grafische Auftragung ist in Abbildung 63 zu finden.



**Abbildung 63:** Variation der Temperatur – Butan MD, Training der Schweratomdistanzmatrix

Auch die Netzkarten aus dem Training der Schweratommatrix lassen erkennen, dass bei höheren Temperaturen vermehrt die *gauche*-Konfiguration in der MD-Simulation vorliegt. Die Qualität der Netzkarten nimmt sowohl subjektiv/visuell in Abbildung 64 als auch objektiv anhand der RMSG- und RMSDSN-Qualitätsparameter in Abbildung 63 in Richtung höherer Temperaturen ab. Dies bestätigt die Ergebnisse aus Kapitel 2.8.1.8, dass die Netzkarten aus Abstandsmatrizen gegenüber thermischem Rauschen nicht robust sind, was die Wahl von Diederwinkeln als Standardtrainingsparameter nahelegt.

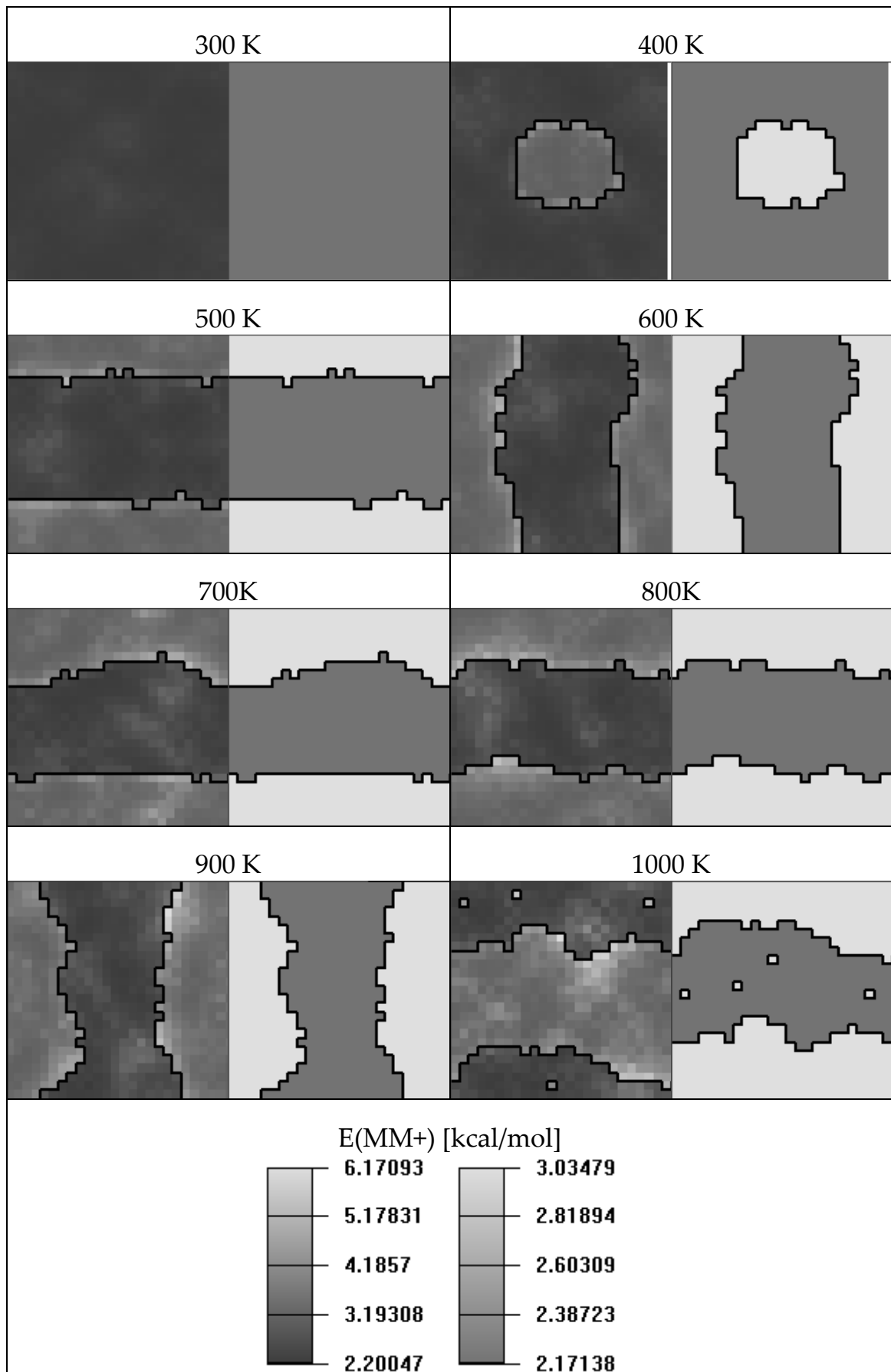
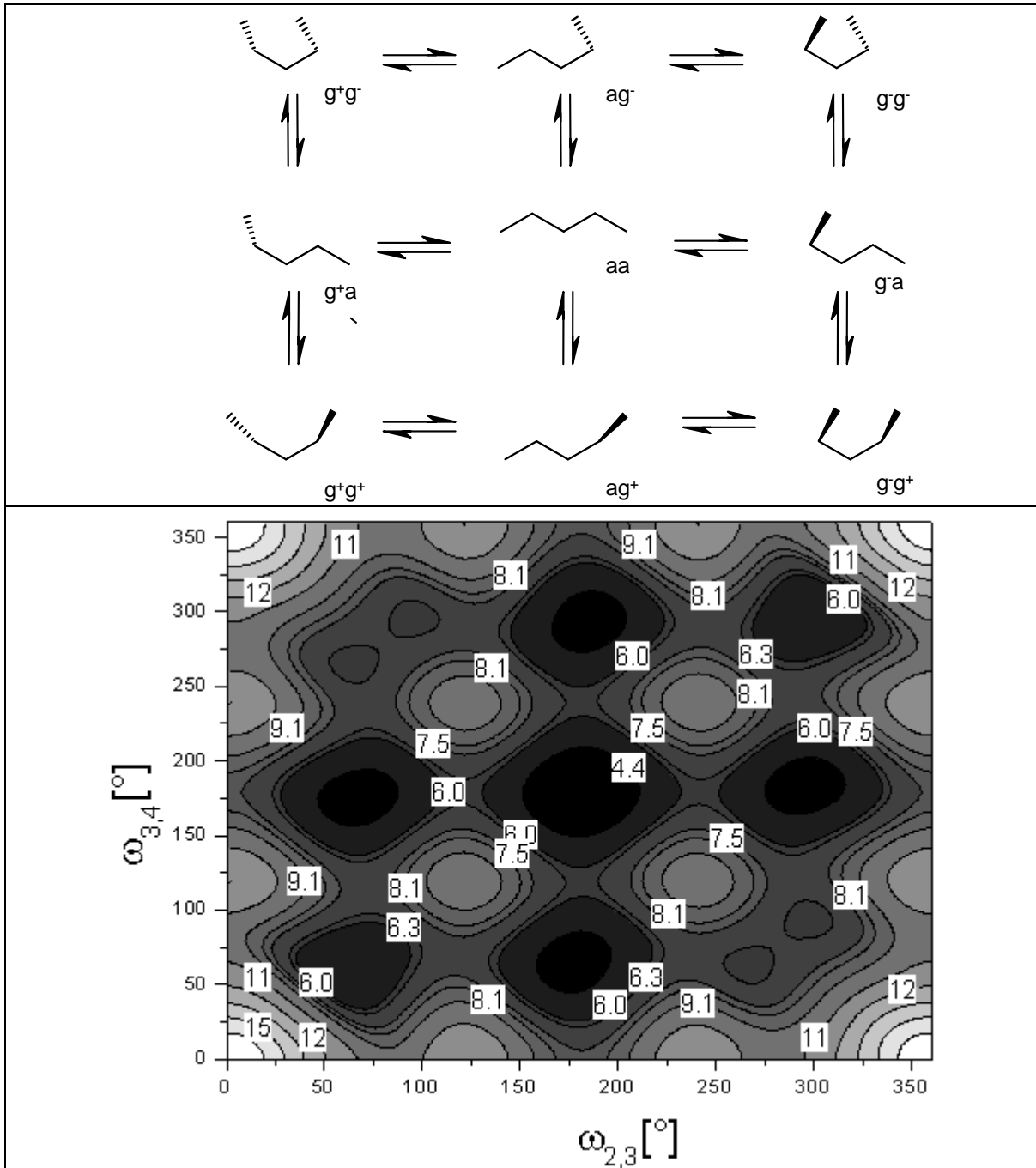


Abbildung 64: Beispielnetzarten Gridhyperfläche – Variation der Temperatur – Butan MD, Training der Schweratomdistanzmatrix, links partiell optimiert nach COLORIX1, rechts vollständig optimiert nach COLORIX2

### 2.8.2 *n*-Pentan

*n*-Pentan kann insgesamt 9 Konformationen einnehmen, wenn man nur die inneren CC-Diederwinkel ohne Stellung der Methylgruppen berücksichtigt, siehe Tabelle 11. Es handelt sich um eine zweidimensionale Hyperfläche von Torustopologie, da für Diederwinkel gilt  $0^\circ \equiv 360^\circ$ , welche ohne Probleme auf die zweidimensionale Netzkarte projiziert werden kann.



**Abbildung 65:** Schematische Darstellung der Pentan Hyperfläche (oben) und zugehörige MM+ Isooberfläche mit Energieisolinien in kcal/mol (unten)

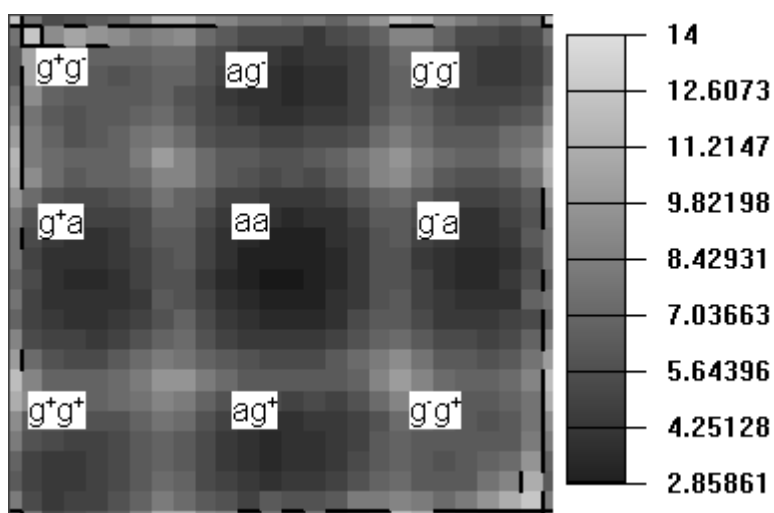
In Abbildung 65 sind die Konformationen nach den Torsionswinkeln benannt ( $a = anti = -120..120^\circ$ ,  $g^+ = gauche^+ = 0..120^\circ$ ,  $g^- = gauche^- = -120..0^\circ$ ). Im oberen Schema sind die konformellen Übergänge mit eingezeichnet. Die MM+ Iso-



berfläche ist das Ergebnis einer *Grid*-Rechnung (Kraftfeld MM+,  $\omega_{2,3} = 0, 5, 10, \dots, 360$  /  $\omega_{3,4} = 0, 5, 10, \dots, 360$ ) über die beiden CC-Diederwinkel des *n*-Pentans mit den Programmen HYPERCHEM und GRID.TCL (5.2.2). Auf der Kraftfeldhyperfläche spalten die  $g^+g^-$  sowie  $g^-g^+$ -Minima aufgrund sterischer Wechselwirkung der endständigen Methylgruppen in zwei sehr flache Nebenminima auf.

### 2.8.2.1 Training der Grid-Hyperfläche

Mit der oben beschriebenen *Grid*-Rechnung wird eine Evaluierung der Netzparameter aus Tabelle 14 durch Training der CC-Diederwinkel analog zum Butanmolekül in Kapitel 2.8.1.5 durchgeführt. Damit ist sichergestellt, dass die vollständige Hyperfläche trainiert wird. Es kann erwartet werden, dass die Hyperfläche aus Abbildung 65 durch Training der beiden Diederwinkel praktisch 1:1 reproduziert werden kann, da sowohl die Hyperfläche des Moleküls als auch die Netzkarte eine torusförmige Topologie besitzen. Projektionsfehler sind in diesem Spezialfall folglich prinzipiell vermeidbar. In der Tat können Netzkarten wie die in Abbildung 66 reproduzierbar erhalten werden.



**Abbildung 66:** Beispielprojektion der MM+ Hyperfläche des Pentans durch Training der beiden CC-Diederwinkel samt Energieskala in kcal/mol

Für die Ermittlung der RMSDCF-Werte wird eine Klassifizierung der Pentan-konformationen mit Hilfe des Programms CLASSIFY\_PENTAN\_SNP.EXE, siehe Kapitel 5.3.4, durchgeführt.

Trainiert werden die Kohonennetze in den folgenden Unterkapiteln mit Vorgabewerten von [ $f_0=0.15$ ,  $epochs=1000$ ,  $neighborhood\ function = parabolisch$ ,  $NETX=NETY=21$ ,  $linear\ learning\ decay$ ] für die nichtvariierten Netzparameter.

### 2.8.2.1.1 Variation der epochs

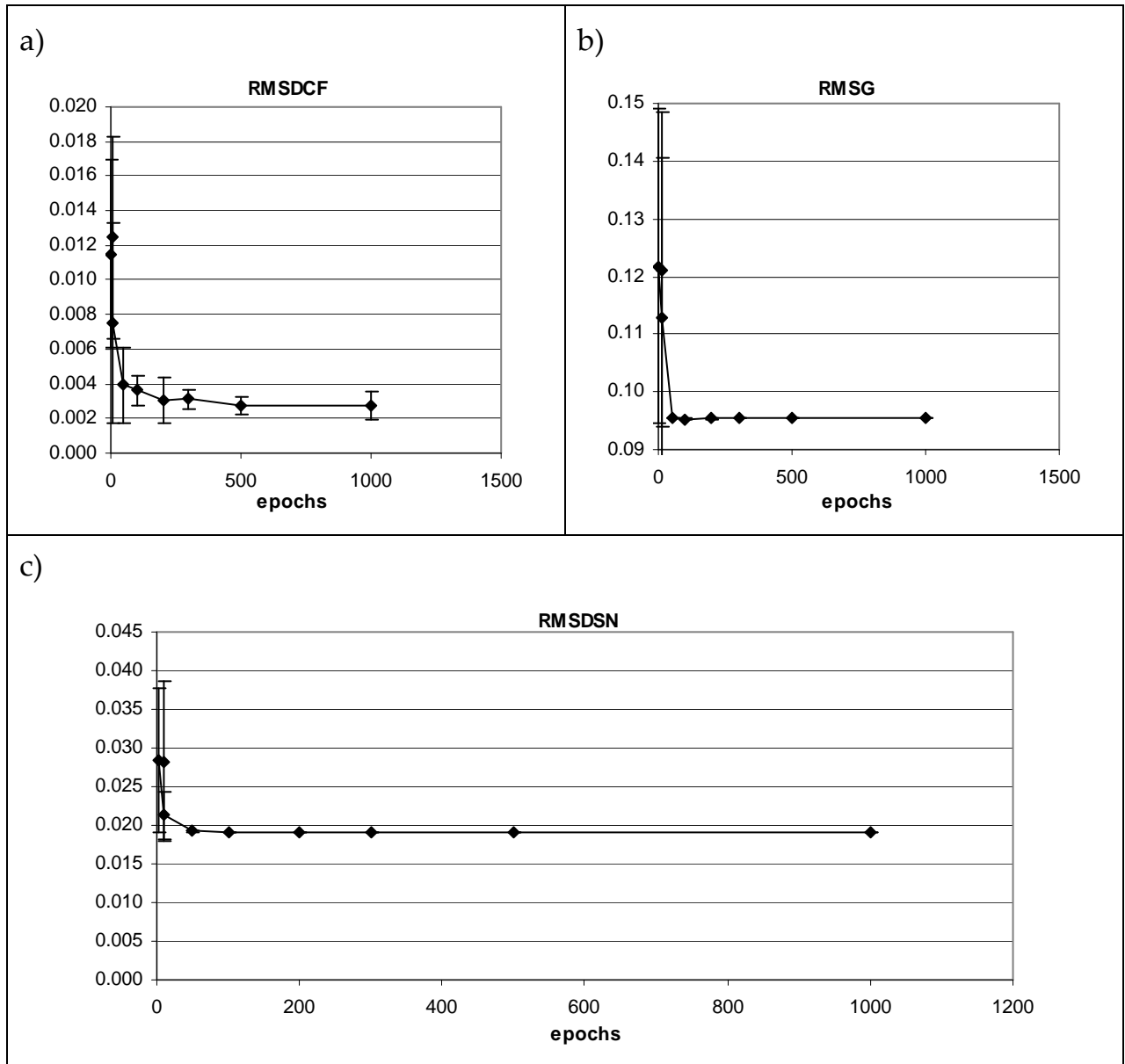
Tabelle 24: Pentan - Variation der epochs

epochs	Ø RMSDCF	$\sigma$	relerr%	min	max	ndata
1000	0.00271	0.00081	29.93	0.00171	0.00389	10
500	0.00271	0.00050	18.59	0.00208	0.00374	10
300	0.00311	0.00058	18.59	0.00251	0.00444	10
200	0.00306	0.00132	43.10	0.00129	0.00525	10
100	0.00361	0.00088	24.48	0.00218	0.00504	10
50	0.00393	0.00216	54.92	0.00121	0.00812	10
10	0.00751	0.00580	77.17	0.00213	0.02007	10
epochs	Ø RMSG	$\sigma$	relerr%	min	max	ndata
1000	0.09534	0.00001	0.01	0.09533	0.09535	10
500	0.09534	0.00001	0.01	0.09533	0.09536	10
300	0.09534	0.00000	0.00	0.09533	0.09534	10
200	0.09533	0.00001	0.01	0.09532	0.09536	10
100	0.09532	0.00001	0.01	0.09531	0.09534	10
50	0.09535	0.00002	0.02	0.09532	0.09537	10
10	0.11276	0.02791	24.76	0.09540	0.15402	10
epochs	Ø RMSDSN	$\sigma$	relerr%	min	max	ndata
1000	0.01910	0.00000	0.01	0.01910	0.01910	10
500	0.01910	0.00000	0.01	0.01910	0.01910	10
300	0.01910	0.00000	0.00	0.01910	0.01910	10
200	0.01911	0.00000	0.02	0.01910	0.01912	10
100	0.01914	0.00001	0.08	0.01913	0.01918	10
50	0.01922	0.00004	0.22	0.01918	0.01930	10
10	0.02127	0.00304	14.28	0.01933	0.02629	10

Die grafische Auftragung obiger Qualitätsparameter ergibt Abbildung 67. Aus den absolut sehr kleinen Werten für RMSG, RMSDCF und RMSDSN mit den extrem kleinen Fehlergrenzen ( $\sigma$ ) kann insgesamt auf sehr hochwertige Projektion geschlossen werden. Des Weiteren zeigen RMSG und RMSDSN an, dass das Netztraining im Rahmen der Fehlergrenzen nach 50 *epochs* bzw. 100 *epochs* beendet ist, so dass die MTS-Werte wie beim Butan abgeschätzt werden.

- $MTS(RMSG)=5184*50=259200$
- $MTS(RMSDSN)=5184*100=518400$

Dies deckt sich größenordnungsmäßig mit den MTS-Werten beim Butan.

Abbildung 67: Variation der *epochs*, a) RMSDCF, b) RMSG, c) RMSDSN

### 2.8.2.1.2 Variation von $f_0$

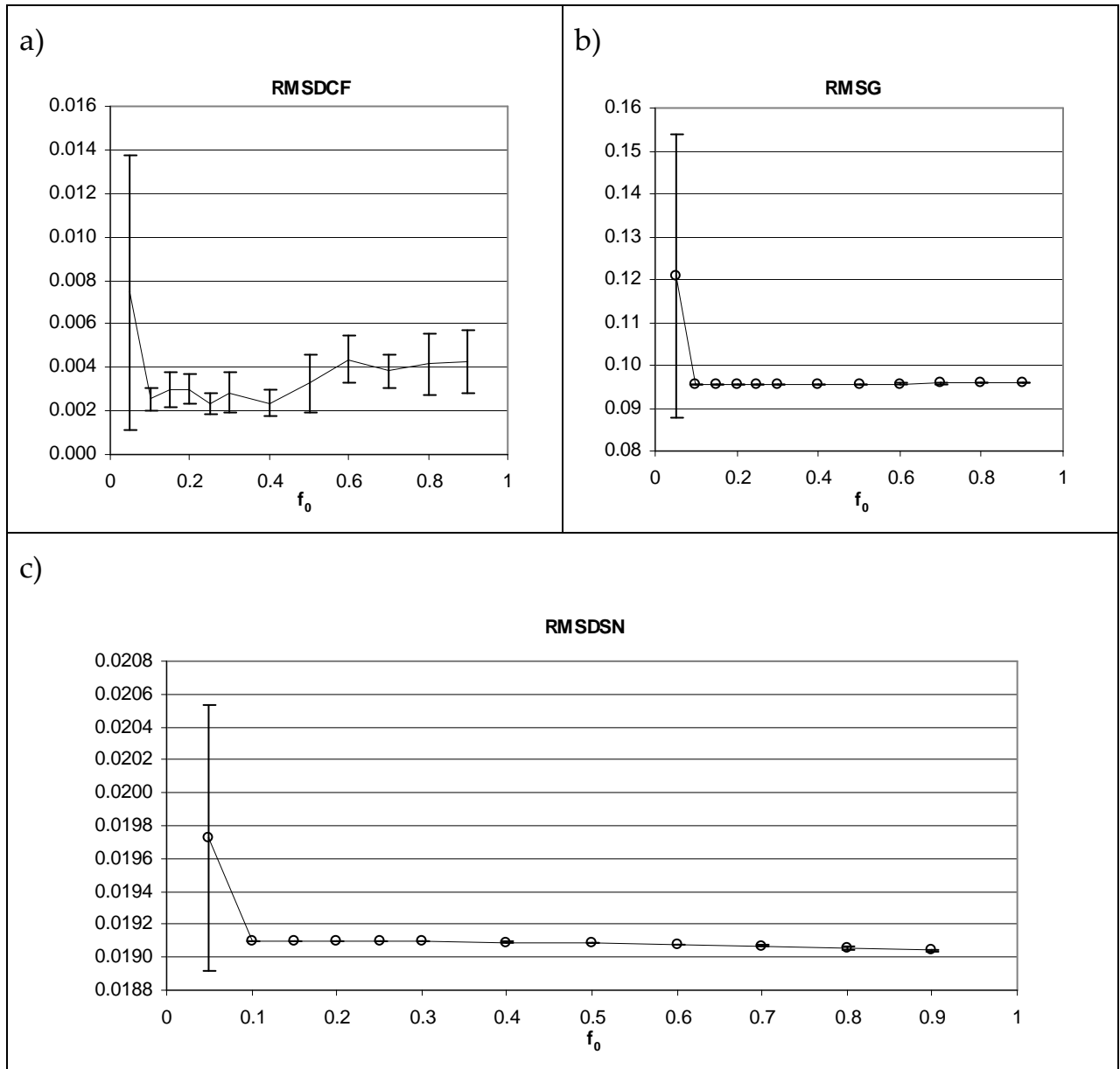
Tabelle 25: Variation des maximalen Adaptionsfaktors  $f_0$  nach Gleichung (2.12)

$f_0$	$\bar{\sigma}$ RMSDCF	$\sigma$	relerr%	Min	max	ndata
0.05	0.00745	0.00631	84.69	0.00249	0.01845	10
0.1	0.00255	0.00050	19.80	0.00205	0.00375	10
0.15	0.00296	0.00083	27.91	0.00208	0.00504	10
0.2	0.00299	0.00069	23.19	0.00188	0.00417	10
0.25	0.00231	0.00050	21.51	0.00149	0.00297	10
0.3	0.00285	0.00090	31.58	0.00183	0.00419	10
0.4	0.00234	0.00060	25.40	0.00149	0.00337	10
0.5	0.00326	0.00130	39.73	0.00174	0.00534	10
0.6	0.00438	0.00111	25.34	0.00265	0.00633	10
0.7	0.00384	0.00076	19.87	0.00293	0.00556	10
0.8	0.00416	0.00139	33.40	0.00154	0.00554	10
0.9	0.00426	0.00148	34.64	0.00231	0.00684	10

$f_0$	Ø RMSG	$\sigma$	relerr%	min	max	ndata
0.05	0.12088	0.03299	27.29	0.09532	0.16019	10
0.1	0.09533	0.00001	0.01	0.09533	0.09534	10
0.15	0.09535	0.00001	0.01	0.09534	0.09536	10
0.2	0.09537	0.00002	0.02	0.09534	0.09539	10
0.25	0.09541	0.00001	0.02	0.09539	0.09544	10
0.3	0.09547	0.00002	0.02	0.09545	0.09551	10
0.4	0.09557	0.00003	0.03	0.09552	0.09561	10
0.5	0.09564	0.00001	0.01	0.09562	0.09565	10
0.6	0.09570	0.00003	0.03	0.09566	0.09576	10
0.7	0.09575	0.00004	0.04	0.09570	0.09581	10
0.8	0.09582	0.00002	0.02	0.09581	0.09585	10
0.9	0.09589	0.00005	0.05	0.09580	0.09594	10
$f_0$	Ø RMSDSN	$\sigma$	relerr%	min	max	ndata
0.05	0.01973	0.00081	4.12	0.01910	0.02069	10
0.1	0.01910	0.00000	0.00	0.01909	0.01910	10
0.15	0.01910	0.00000	0.00	0.01909	0.01910	10
0.2	0.01910	0.00000	0.01	0.01909	0.01910	10
0.25	0.01910	0.00000	0.01	0.01910	0.01910	10
0.3	0.01910	0.00000	0.01	0.01910	0.01910	10
0.4	0.01909	0.00000	0.02	0.01908	0.01910	10
0.5	0.01909	0.00000	0.01	0.01908	0.01909	10
0.6	0.01908	0.00000	0.02	0.01907	0.01908	10
0.7	0.01907	0.00000	0.02	0.01906	0.01908	10
0.8	0.01906	0.00001	0.04	0.01905	0.01907	10
0.9	0.01904	0.00001	0.04	0.01903	0.01906	10

Eine grafische Auswertung der Tabelle 25 ist in Abbildung 68 zu finden. Es bestätigt sich die Beobachtung aus dem Butankapitel 2.8.1.5.2, dass die Qualität der erhaltenen Netzkarten, ausgedrückt durch die RMSG- und RMSDSN-Parameter, bei genügend langem Training über weite Bereiche weitgehend unabhängig von den  $f_0$ -Werten ist. Allerdings kann in diesem Fall ein verschlechterter RMSDCF-Wert, für Werte von  $f_0 > 0.4$  nachgewiesen werden.

Als Trend kann außerdem festgehalten werden, dass die RMSG-Werte bei ca. 0.1 .. 0.15 ein Optimum haben, während die RMSDSN-Werte generell besser werden für größere  $f_0$ .

Abbildung 68: Variation von  $f_0$ , a) RMSDCF, b) RMSG, c) RMSDSN

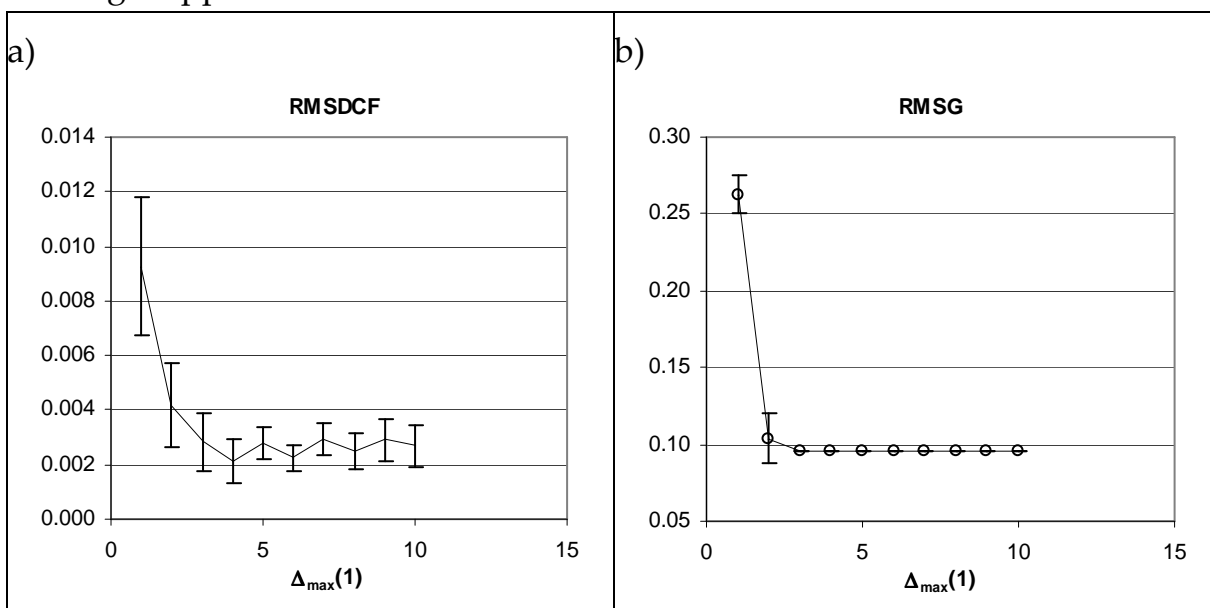
### 2.8.2.1.3 Variation von $\Delta_{\max}(1)$

Tabelle 26: Variation von  $\Delta_{\max}(1)$  nach Gleichung (2.15)

$\Delta_{\max}(1)$	$\bar{\sigma}$ RMSDCF	$\sigma$	relerr%	min	max	ndata
10	0.00268	0.00075	28.01	0.00121	0.00392	10
9	0.00291	0.00074	25.62	0.00208	0.00446	10
8	0.00249	0.00069	27.83	0.00149	0.00422	10
7	0.00292	0.00061	20.76	0.00208	0.00383	10
6	0.00224	0.00049	21.75	0.00132	0.00291	10
5	0.00280	0.00059	20.92	0.00208	0.00378	10
4	0.00212	0.00081	38.11	0.00078	0.00365	10
3	0.00284	0.00106	37.26	0.00151	0.00524	10
2	0.00415	0.00154	37.05	0.00229	0.00795	10
1	0.00927	0.00254	27.34	0.00595	0.01320	10

$\Delta_{\max}(1)$	$\bar{\sigma}$ RMSG	$\sigma$	relerr%	min	max	ndata
10	0.09534	0.00001	0.01	0.09533	0.09535	10
9	0.09536	0.00001	0.01	0.09534	0.09537	10
8	0.09536	0.00001	0.01	0.09535	0.09537	10
7	0.09543	0.00002	0.02	0.09540	0.09545	10
6	0.09548	0.00002	0.02	0.09544	0.09551	10
5	0.09546	0.00001	0.01	0.09544	0.09547	10
4	0.09555	0.00002	0.02	0.09551	0.09558	10
3	0.09566	0.00003	0.03	0.09560	0.09572	10
2	0.10362	0.01634	15.77	0.09579	0.13529	10
1	0.26265	0.01202	4.58	0.25150	0.28410	10
$\Delta_{\max}(1)$	$\bar{\sigma}$ RMSDSN	$\sigma$	relerr%	min	max	ndata
10	0.01910	0.00000	0.01	0.01909	0.01910	10
9	0.01910	0.00000	0.01	0.01909	0.01910	10
8	0.01910	0.00000	0.00	0.01910	0.01910	10
7	0.01910	0.00000	0.00	0.01910	0.01910	10
6	0.01910	0.00000	0.01	0.01909	0.01910	10
5	0.01910	0.00000	0.01	0.01909	0.01910	10
4	0.01909	0.00000	0.02	0.01909	0.01910	10
3	0.01908	0.00000	0.02	0.01908	0.01909	10
2	0.01908	0.00013	0.69	0.01901	0.01945	10
1	0.01974	0.00010	0.50	0.01963	0.01996	10

Die grafische Auswertung ergibt ähnlich wie beim Butan eine „untere Schranke“ für  $\Delta_{\max}(1)$ , in diesem Fall bei  $\Delta_{\max}(1)=3$  (30%). Als Arbeitshypothese kann an dieser Stelle angenommen werden, dass diese Schranke niedriger liegt, je mehr *epochs* trainiert werden. Sowohl mit Verringerung der *epochs* als auch durch Verringerung von  $\Delta_{\max}(1)$  lässt sich Rechenzeit sparen, allerdings ist zu erwarten, dass die minimalen Größen für beide Netzparameter stark miteinander gekoppelt sind



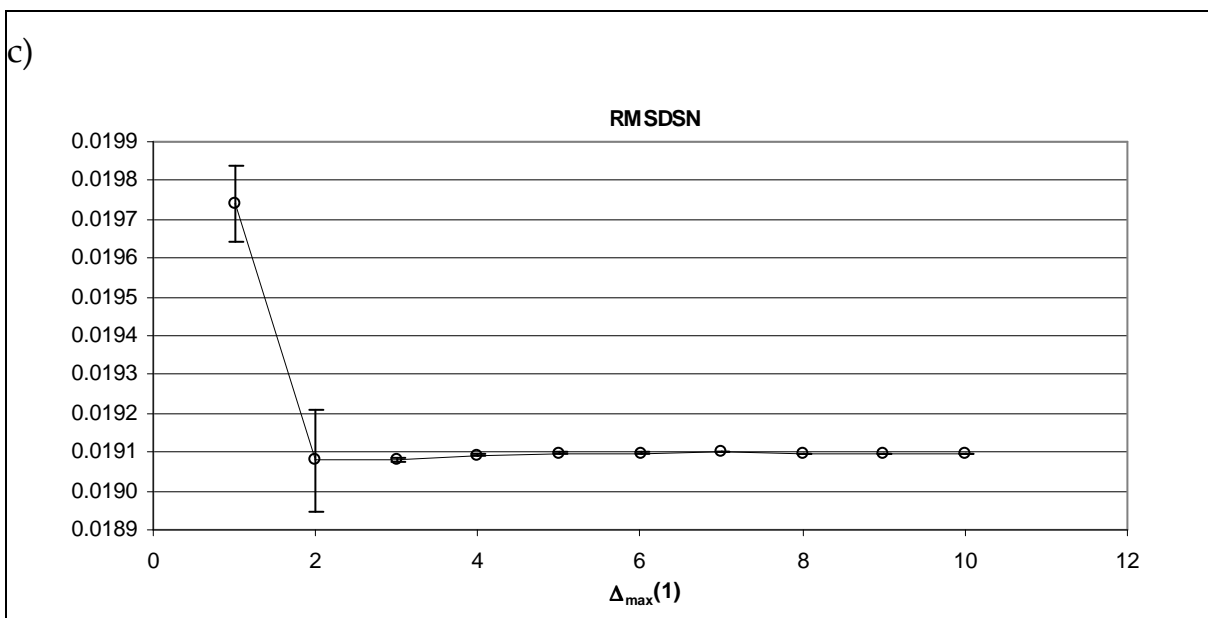


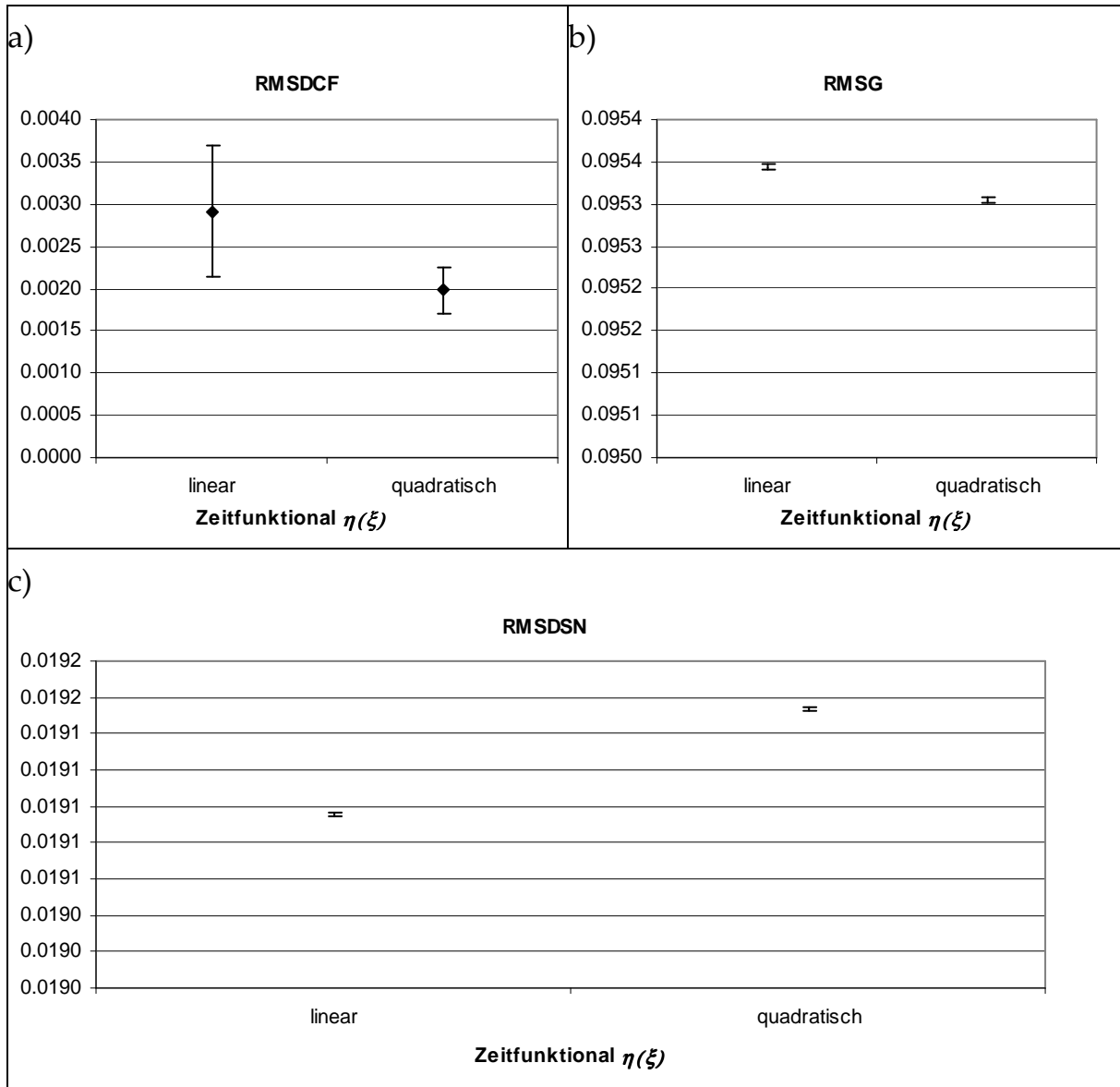
Abbildung 69: Variation von  $\Delta_{\max}(1)$ , a) RMSDCF, b) RMSG, c) RMSDSN

### 2.8.2.1.4 Variation von $\eta(\xi)$

Tabelle 27: Variation der Zeitabhängigkeit des Adaptionfaktors,  $\eta(\xi)$  nach Gleichung (2.17)

$\eta(\xi)$ (2.17)	$\bar{\Delta}$ RMSDCF	$\sigma$	relerr%	min	max	ndata
a) linear	0.00292	0.00078	26.9	0.00210	0.00482	10
b) quadratisch	0.00198	0.00026	13.4	0.00161	0.00237	10
$\eta(\xi)$ (2.17)	$\bar{\Delta}$ RMSG	$\sigma$	relerr%	min	max	ndata
a) linear	0.09534	0.000003	0.0	0.09534	0.09535	10
b) quadratisch	0.09530	0.000003	0.0	0.09530	0.09531	10
$\eta(\xi)$ (2.17)	$\bar{\Delta}$ RMSDSN	$\sigma$	relerr%	min	max	ndata
a) linear	0.01910	0.000001	0.0	0.01909	0.01910	10
b) quadratisch	0.01915	0.000001	0.0	0.01915	0.01916	10

Wie beim Butan zeigt sich der marginale Einfluss der Wahl des Adaptionfaktors  $\eta(\xi)$ . Während beim RMSDSN-Wert wie beim Butan leichte Vorteile der linearen Funktion zu erkennen sind, zeigen die RMSDCF- und RMSG-Werte Vorteile für die quadratische Funktion.

Abbildung 70: Variation von  $\eta(\xi)$  nach Gleichung (2.17)

### 2.8.2.1.5 Variation von NETX, NETY

Tabelle 28: Variation der Netzgröße NETX=NETY

NETX=NETY	Ø RMSDCF	$\sigma$	relerr%	min	max	ndata
31	0.00357	0.00051	14.36	0.00267	0.00405	10
21	0.00267	0.00046	17.33	0.00208	0.00357	10
11	0.01840	0.00000	0.00	0.01840	0.01840	10
5	0.03735	0.00595	15.92	0.02627	0.04731	10
NETX=NETY	Ø RMSG	$\sigma$	relerr%	min	max	ndata
31	0.06487	0.00001	0.02	0.06485	0.06488	10
21	0.09534	0.00001	0.01	0.09533	0.09535	10
11	0.18205	0.00002	0.01	0.18201	0.18209	10
5	0.40832	0.02483	6.08	0.40038	0.47899	10
NETX=NETY	Ø RMSDSN	$\sigma$	relerr%	min	max	ndata
31	0.06487	0.00001	0.02	0.06485	0.06488	10
21	0.09534	0.00001	0.01	0.09533	0.09535	10
11	0.18205	0.00002	0.01	0.18201	0.18209	10
5	0.40832	0.02483	6.08	0.40038	0.47899	10



Die grafische Auswertung der Daten aus der Variation der Netzgröße in Abbildung 71 zeigt für die RMSG- und RMSDSN-Werte wieder höhere Qualität bei größeren Netzkarten. Interessant ist jedoch der RMSDCF-Wert, der bei ausnahmsweise äußerst geringen Fehlergrenzen ein Optimum für eine Netzgröße von  $\text{NETX}=\text{NETY}=21$  zeigt. Insgesamt bestätigen sich die Ergebnisse, welche beim Butan gewonnen wurden.

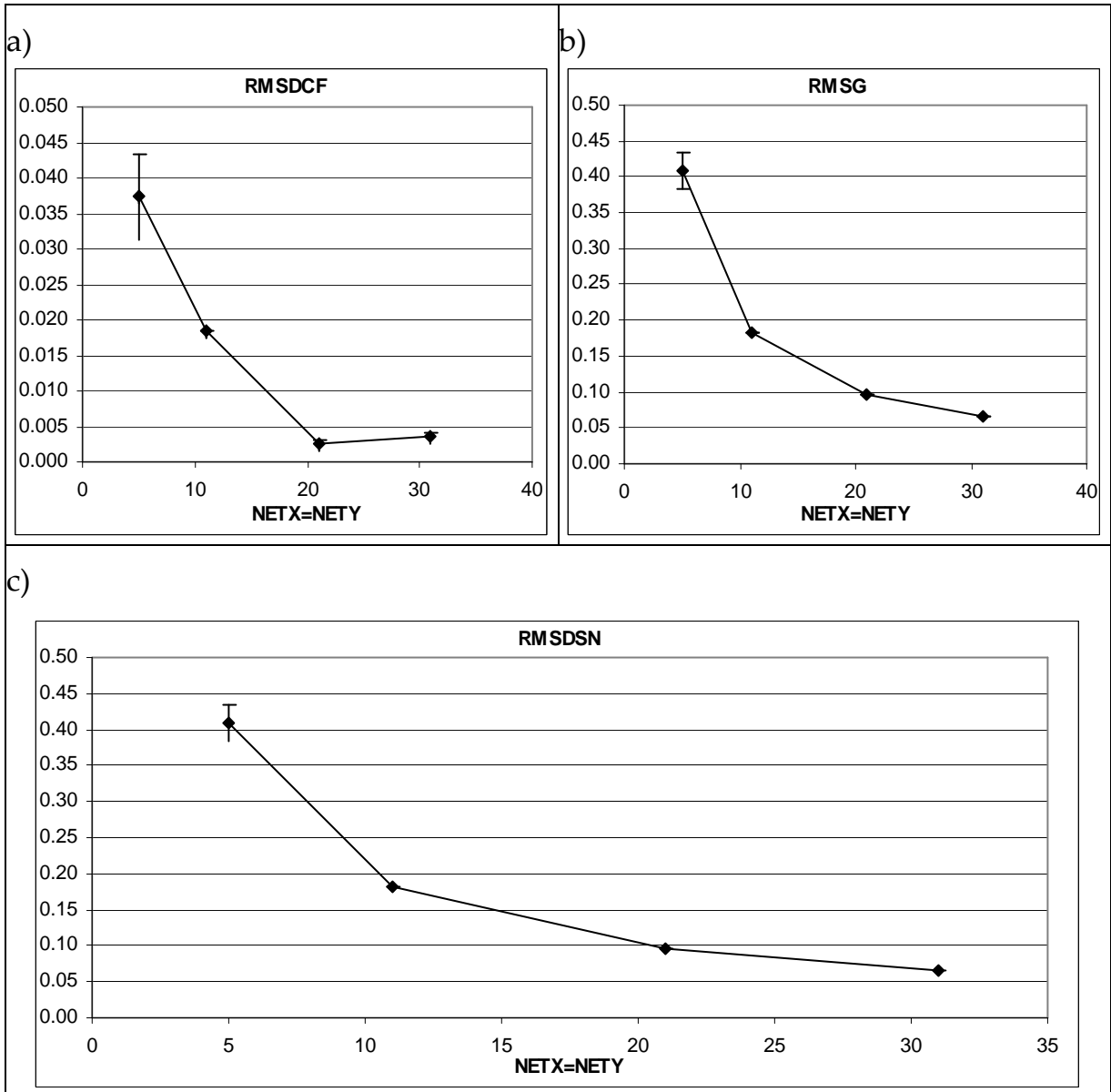


Abbildung 71: Variation der Netzgröße  $\text{NETX}=\text{NETY}$

### 2.8.2.1.6 Variation der Abstandsfunktion $\vartheta(\zeta)$

**Tabelle 29:** Variation der Abstandsfunktion  $\vartheta(\zeta)$  nach Gleichung (2.18)

$\vartheta(\zeta)$	$\emptyset$ RMSDCF	$\sigma$	re- lerr%	min	max	nda- ta
<i>block</i>	0.00379	0.00065	17.12	0.00293	0.00520	10
<i>Mexican hat</i>	0.01031	0.00256	24.78	0.00752	0.01611	10
parabolisch	0.00248	0.00033	13.21	0.00205	0.00297	10
<i>pseudo Mexican</i>	0.00594	0.00395	66.56	0.00308	0.01648	10
quadratisch- parabolisch	0.00290	0.00049	16.88	0.00205	0.00392	10
quadratisch- triangular	0.00301	0.00061	20.38	0.00210	0.00405	10
triangular	0.00285	0.00039	13.76	0.00251	0.00346	10
$\vartheta(\zeta)$	$\emptyset$ RMSG	$\sigma$	re- lerr%	min	max	nda- ta
<i>block</i>	0.09537	0.00000	0.00	0.09537	0.09538	10
<i>Mexican hat</i>	0.74962	0.01963	2.62	0.72120	0.78132	10
parabolisch	0.09534	0.00001	0.01	0.09534	0.09536	10
<i>pseudo Mexican</i>	0.49190	0.03825	7.78	0.47639	0.59984	10
quadratisch- parabolisch	0.09538	0.00001	0.01	0.09536	0.09539	10
quadratisch- triangular	0.09537	0.00001	0.01	0.09536	0.09538	10
triangular	0.09537	0.00001	0.01	0.09536	0.09540	10
$\vartheta(\zeta)$	$\emptyset$ RMSDSN	$\sigma$	re- lerr%	min	max	nda- ta
<i>block</i>	0.01928	0.00000	0.02	0.01927	0.01928	10
<i>Mexican hat</i>	0.02225	0.00051	2.28	0.02171	0.02313	10
parabolisch	0.01910	0.00000	0.01	0.01909	0.01910	10
<i>pseudo Mexican</i>	0.01963	0.00060	3.03	0.01929	0.02130	10
quadratisch- parabolisch	0.01909	0.00000	0.00	0.01909	0.01909	10
quadratisch- triangular	0.01909	0.00000	0.00	0.01909	0.01909	10
triangular	0.01909	0.00000	0.01	0.01909	0.01909	10

Die grafische Auswertung, Abbildung 72, der Versuchsreihe in Tabelle 29 bestätigt das Ergebnis der Butan-Versuchsreihe. Die „*Mexican hat*“-artigen Funktionen sind ungeeignet als Abstandsfunktion. Ansonsten gibt es keine großen Unterschiede, wobei die „parabolische“ Funktion summa summarum

wahrscheinlich am besten abschneidet. Der hohe RMSDSN-Wert der *block*-Funktion beim Butan wird für das Pentan nicht bestätigt.

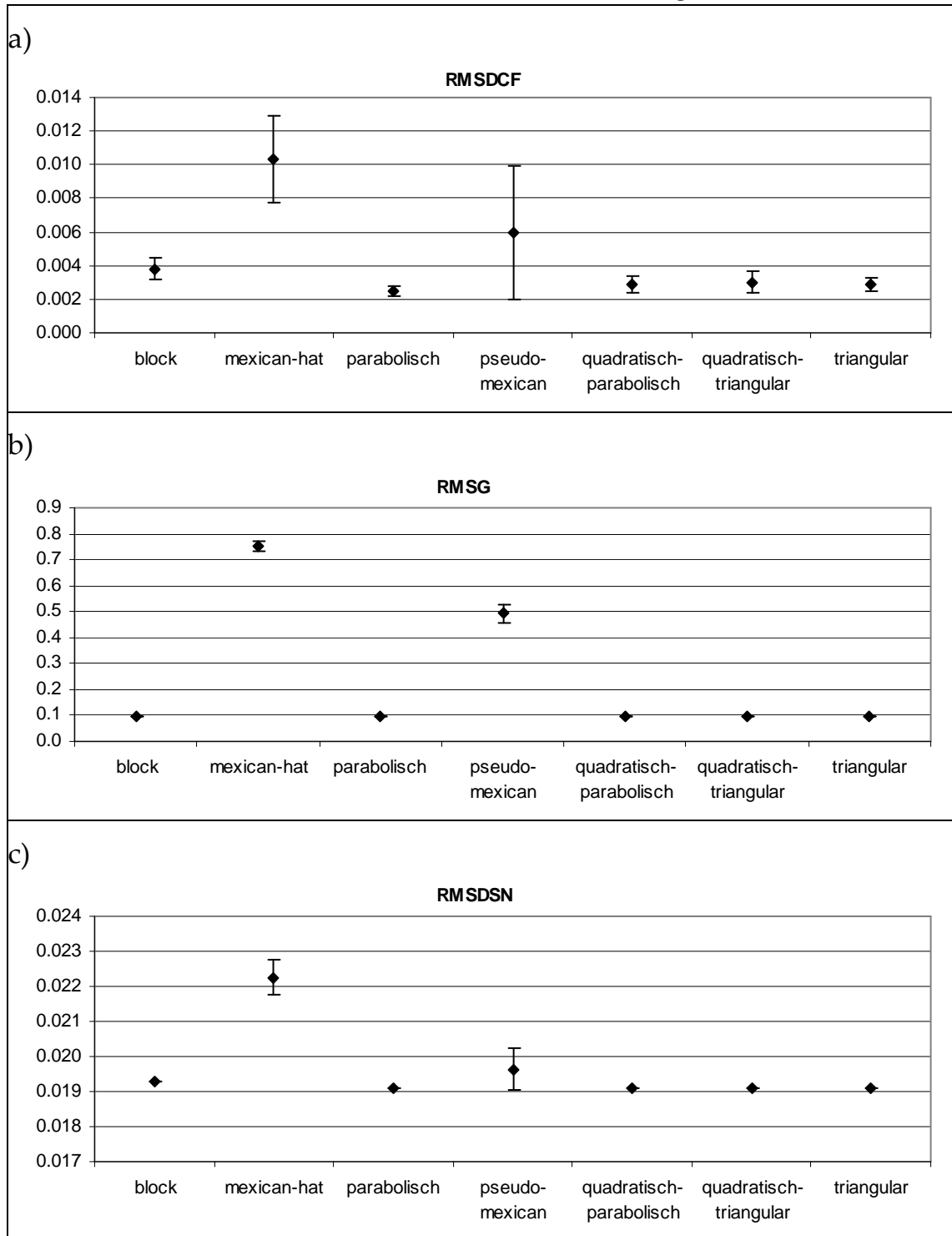


Abbildung 72: Variation der Abstandsfunktion  $\vartheta(\zeta)$

### 2.8.2.1.7 Fazit

Es bestätigen sich weitgehend die Ergebnisse aus 2.8.1.5.7, allerdings mit leicht variierten Optimumswerten.

- *epochs*: ermittelte minimale Trainingsschritte nach Kapitel 2.5.4  
 MTS(RMSG)=259200  
 MTS(RMSDSN)=518400
- $f_0$ :  
 Nach RMSG und RMSDCF scheint ein Wert von ca. 0.10..0.15 ein Optimum darzustellen, was dem Standard-Wert der Trainingsreihe entspricht.
- $\Delta_{\max}(1)$   
 Der RMSG-Wert bricht ab ca. 5 Neuronen (50%) ein. Der Standardwert von 10 Neuronen (100%) ist definitiv höher als nötig.
- $\eta(\xi)$   
 Die lineare Funktion ist besser geeignet, um einen niedrigen RMSDSN-Wert zu erhalten, die quadratische Funktion („*quadratic learning decay=on*“) besser für einen niedrigen RMSG-Wert.
- NETX / NETY  
 Ein Fehler > 5% für RMSDSN und RMSG ergibt sich in obiger Analyse bei  $\leq 25$  Neuronen, was 3 Neuronen/Konformation entspricht.
- $\vartheta(\zeta)$   
 Weitgehend identisch zur Butanversuchsreihe, die „parabolische“ Abstandsfunktion bewährt sich als Standard.

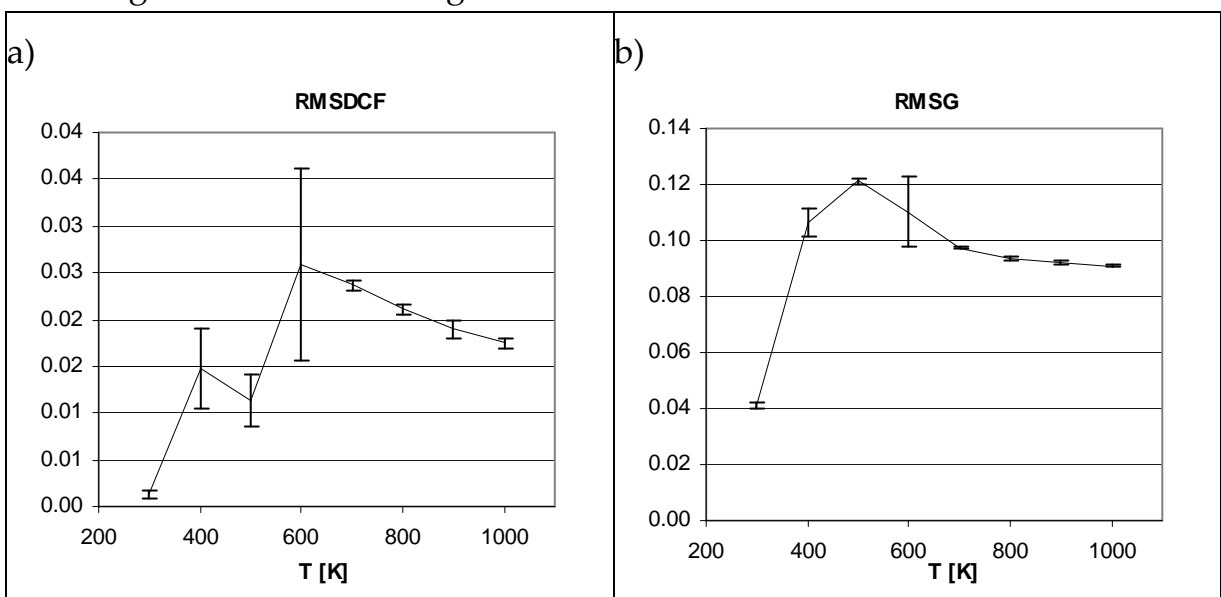
### 2.8.2.2 MD Pentan, Training der Diederwinkel

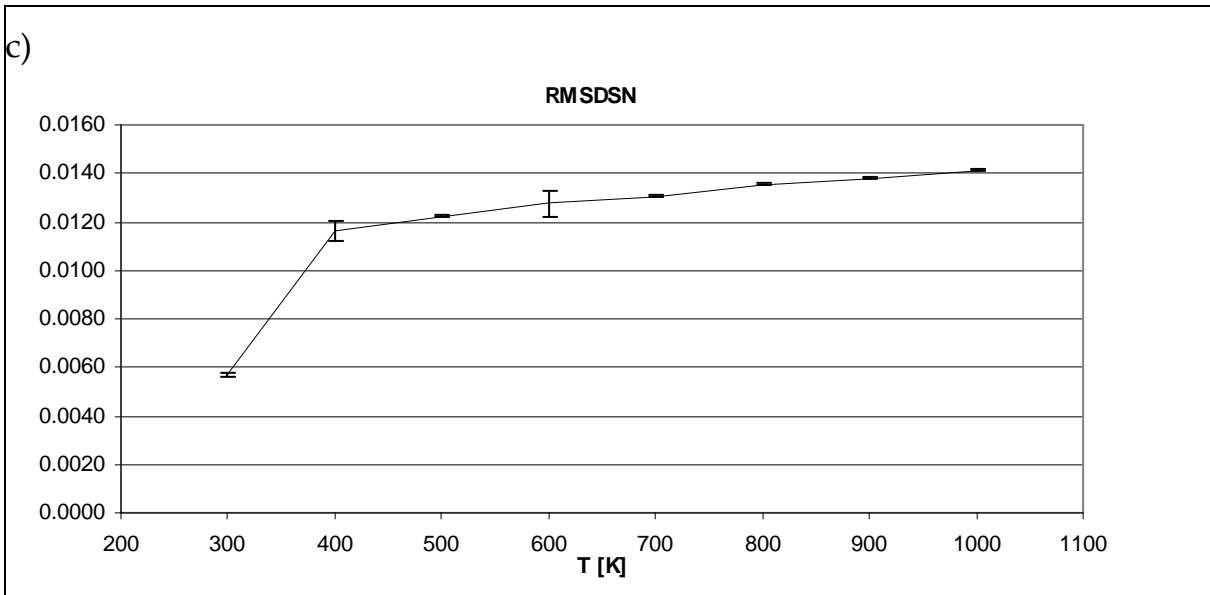
Genau wie für das Butan wird im Folgenden eine Gruppe Moleküldynamik-rechnungen trainiert, welche bei unterschiedlichen Temperaturen simuliert wurden (Kraftfeld HYPERCHEM: MM+, in vacuo, 20000 Strukturen, Simulationsdauer 10000 ps, Aufruf von FIX\_ROTATION.EXE bei je 1 ps, *step size* 0.001 ps, *datafreq*=1/500 *data steps*<sup>-1</sup>). Die erhaltenen Daten sind in Tabelle 30 angeführt. Die Netzparameter betragen: [25x25 Neuronen; 100 *epochs*;  $\Delta_{\max}(1)$ =12;  $f_0$ =0.4; Abstandsfunktion parabolisch, Training der beiden CC-Diederwinkel].

**Tabelle 30: Pentan MD-Simulationen bei unterschiedlichen Temperaturen, Training der zentralen Diederwinkel**

T [K]	Ø RMSDCF	$\sigma$	relerr%	min	max	ndata
1000	0.01744	0.00045	2.61	0.01649	0.01796	10
900	0.01894	0.00093	4.92	0.01747	0.02009	10
800	0.02112	0.00053	2.49	0.02047	0.02201	10
700	0.02366	0.00059	2.50	0.02278	0.02492	10
600	0.02594	0.01029	39.66	0.00619	0.03319	10
500	0.01135	0.00277	24.39	0.00669	0.01574	10
400	0.01479	0.00433	29.26	0.00741	0.02006	10
300	0.00127	0.00043	33.46	0.00060	0.00210	10
T [K]	Ø RMSG	$\sigma$	relerr%	min	max	ndata
1000	0.09090	0.00051	0.57	0.09016	0.09174	10
900	0.09233	0.00064	0.69	0.09145	0.09312	10
800	0.09371	0.00051	0.54	0.09322	0.09442	10
700	0.09728	0.00044	0.45	0.09636	0.09768	10
600	0.11031	0.01269	11.50	0.10022	0.12700	10
500	0.12116	0.00110	0.91	0.11885	0.12256	10
400	0.10645	0.00494	4.64	0.10034	0.11301	10
300	0.04105	0.00092	2.24	0.03976	0.04266	10
T [K]	Ø RMSDSN	$\sigma$	relerr%	min	max	ndata
1000	0.01412	0.00004	0.28	0.01406	0.01418	10
900	0.01378	0.00004	0.32	0.01371	0.01386	10
800	0.01355	0.00002	0.18	0.01351	0.01359	10
700	0.01306	0.00004	0.32	0.01298	0.01312	10
600	0.01275	0.00050	3.92	0.01236	0.01370	10
500	0.01223	0.00005	0.44	0.01213	0.01230	10
400	0.01163	0.00043	3.73	0.01081	0.01194	10
300	0.00570	0.00010	1.76	0.00554	0.00583	10

Die absoluten Werte der Qualitätskriterien sind größenordnungsmäßig vergleichbar mit denen der synthetischen Hyperfläche. Dargestellt ist obige Tabelle grafisch in Abbildung 73.





**Abbildung 73: Variation der Temperatur – Pentan MD, Training der beiden CC-Diederwinkel**

In der grafischen Auswertung, Abbildung 73, ergibt sich ein interessanter Effekt des Ansteigens der RMSG- und RMSDCF-Werte zwischen 300 und 600 K. Dies wird bewirkt durch die steigende Zahl der zu trainierenden höherenergetischen Konformationen. Ab 700 K kann die Hyperfläche dann als vollständig abgebildet betrachtet werden, mit asymptotischer Verbesserung für höhere Temperaturen. Die RMSDSN-Werte steigen kontinuierlich an, so dass deren Verwendung als Maßstab der Abbildungsqualität hier problematisch ist. Das schrittweise Auftreten der höherenergetischen Konformationen in den Simulationen kann durch visuelle Inspektion der Abbildung 74 einfach nachvollzogen werden.

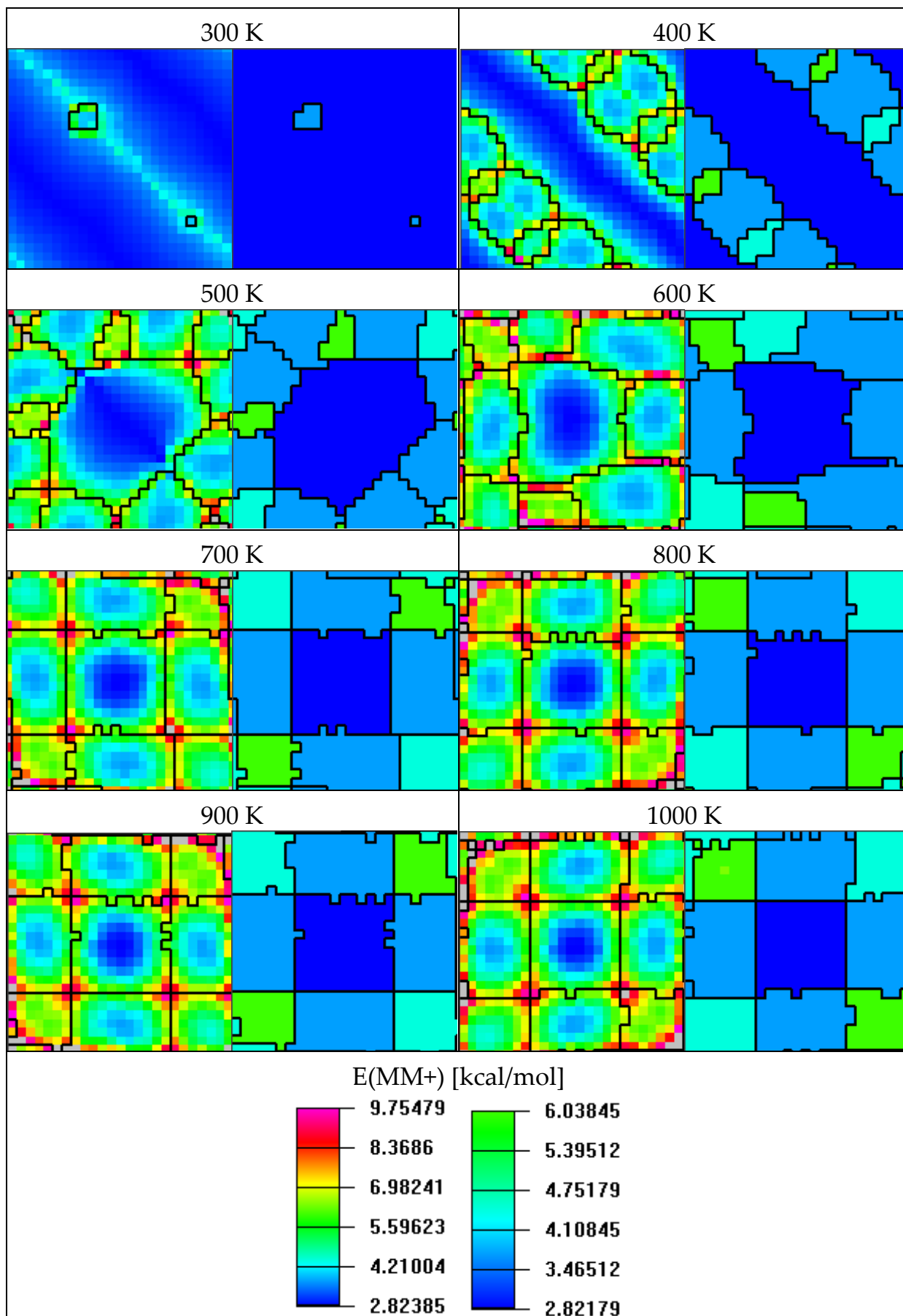


Abbildung 74: Beispielnetskarten Gridhyperfläche – Variation der Temperatur – Pentan MD, Training der zentralen Diederwinkel, links partiell optimiert unter Verwendung der Netzrestraints nach COLORIX1, rechts vollständig optimiert nach COLORIX2

### 2.8.2.3 MD Pentan, Training aller Distanzen

Das Training der C-C, C-H und H-H Abstände ergibt wie beim Butan eine komplexe Hyperfläche mit Numerationsisomeren. Aus Rechenzeitgründen werden nur 20 *epochs* und je 5 Wiederholungen jeder Rechnung durchgeführt. Es zeigt sich bereits bei diesem relativ kleinen Molekül, dass die komplette Distanzmatrix zu rechenzeitintensiv wird. Die statistischen Daten des Netztrainings der Moleküldynamiken aus Kapitel 2.8.2.2 sind in Tabelle 31 aufgeführt. Als Netzparameter wurden gewählt: [25x25 Neuronen; 20 *epochs*;  $\Delta_{\max}(1)=12$ ;  $f_0=0.4$ ; Abstandsfunktion parabolisch, Training aller Abstände].

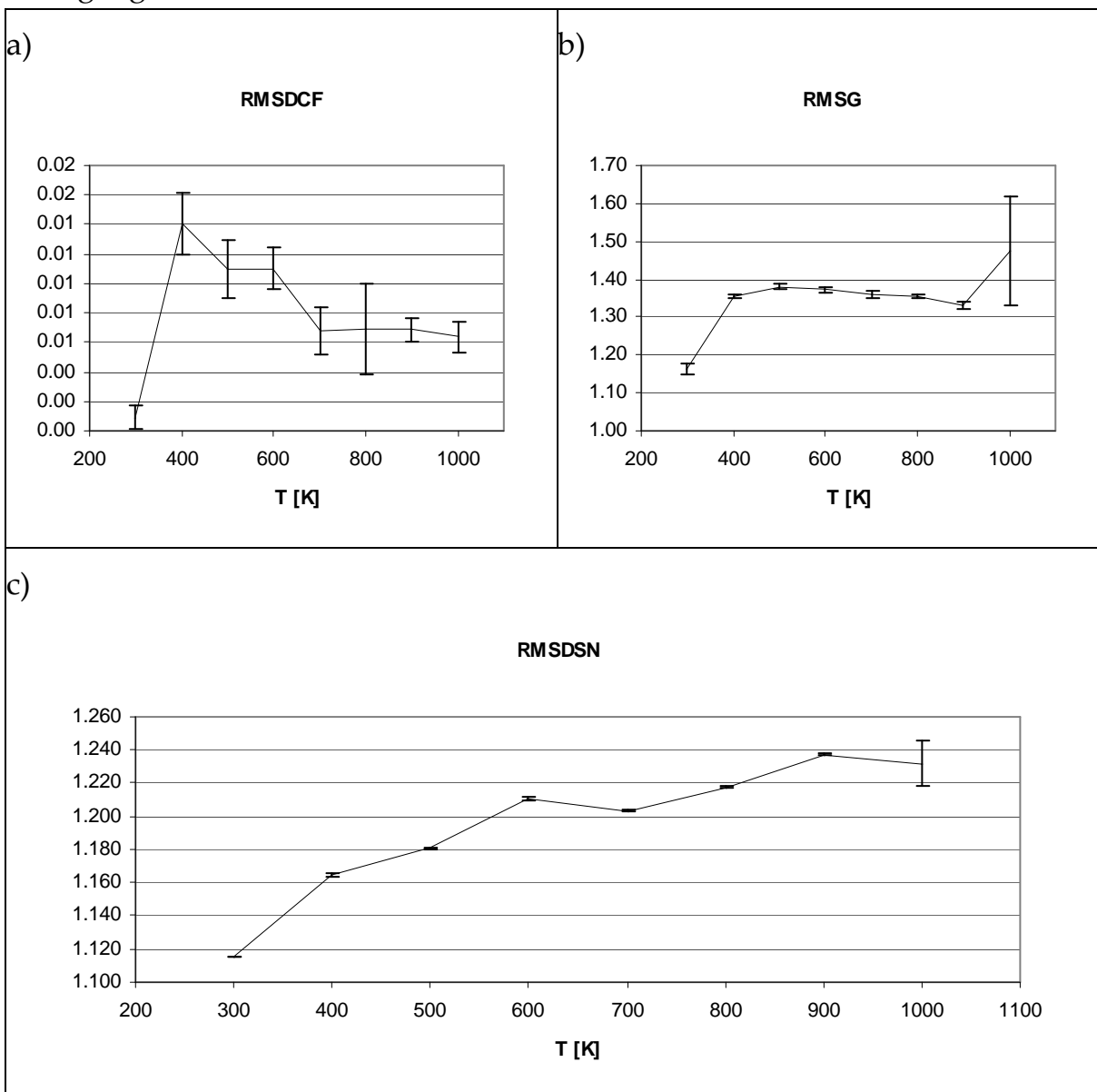
**Tabelle 31: Pentan MD-Simulationen bei unterschiedlichen Temperaturen, Training der kompletten Distanzmatrix**

T [K]	Ø RMSDCF	$\sigma$	relerr%	min	max	ndata
1000	0.00637	0.00105	16.44	0.00493	0.00748	5
900	0.00686	0.00077	11.27	0.00561	0.00741	5
800	0.00695	0.00307	44.14	0.00374	0.01163	5
700	0.00676	0.00159	23.52	0.00482	0.00824	5
600	0.01102	0.00144	13.03	0.00947	0.01239	5
500	0.01094	0.00199	18.15	0.00879	0.01356	5
400	0.01406	0.00205	14.59	0.01091	0.01593	5
300	0.00089	0.00083	92.97	0.00020	0.00210	5
T [K]	Ø RMSG	$\sigma$	relerr%	min	max	ndata
1000	1.47419	0.14535	9.86	1.30848	1.58737	5
900	1.33131	0.00894	0.67	1.32011	1.34086	5
800	1.35383	0.00380	0.28	1.34834	1.35812	5
700	1.35760	0.00942	0.69	1.34705	1.37262	5
600	1.37234	0.00645	0.47	1.36537	1.38107	5
500	1.37964	0.00806	0.58	1.37068	1.38871	5
400	1.35350	0.00402	0.30	1.34988	1.35944	5
300	1.16232	0.01326	1.14	1.14684	1.17728	5
T [K]	Ø RMSDSN	$\sigma$	relerr%	min	max	ndata
1000	1.23203	0.01354	1.10	1.22147	1.24725	5
900	1.23725	0.00069	0.06	1.23614	1.23794	5
800	1.21738	0.00062	0.05	1.21674	1.21838	5
700	1.20341	0.00043	0.04	1.20297	1.20403	5
600	1.21098	0.00098	0.08	1.21016	1.21261	5
500	1.18063	0.00101	0.09	1.17927	1.18156	5
400	1.16489	0.00081	0.07	1.16384	1.16605	5
300	1.11526	0.00043	0.04	1.11467	1.11579	5

Die grafische Auftragung der Qualitätsparameter ist in Abbildung 75 zu sehen. Bei 300 K ist offensichtlich nur ein kleiner Teil der Hyperfläche abgebildet, siehe die absolut niedrigen RMSDCF- und RMSG-Werte. Der RMSDCF-Wert sinkt ab 400 K kontinuierlich ab, d.h. bei höheren Temperaturen sind die Flächenverhältnisse der Netzkarte repräsentativer für die Anteile der abgebildeten Konformationen an der Trajektorie. Genau wie beim Training der



Diederwinkel zeigt sich ein RMSG-Maximum bei 500 K, welches allerdings weniger deutlich ausgebildet ist. Bei dieser Temperatur unterliegen die Trainingsdaten anscheinend einer Zusammensetzung, welche beim Training einen hohen Netzgradienten zur Folge hat. Dies ist in der Regel mit dem Übergang von einer Netzkonformation in eine andere verbunden.



**Abbildung 75: Variation der Temperatur – Pentan MD, Training der kompletten Distanzmatrix**

Die Abbildung 76 zeigt beispielhaft Netzkartern, wie sie durch das Training aller Distanzen erhalten werden.

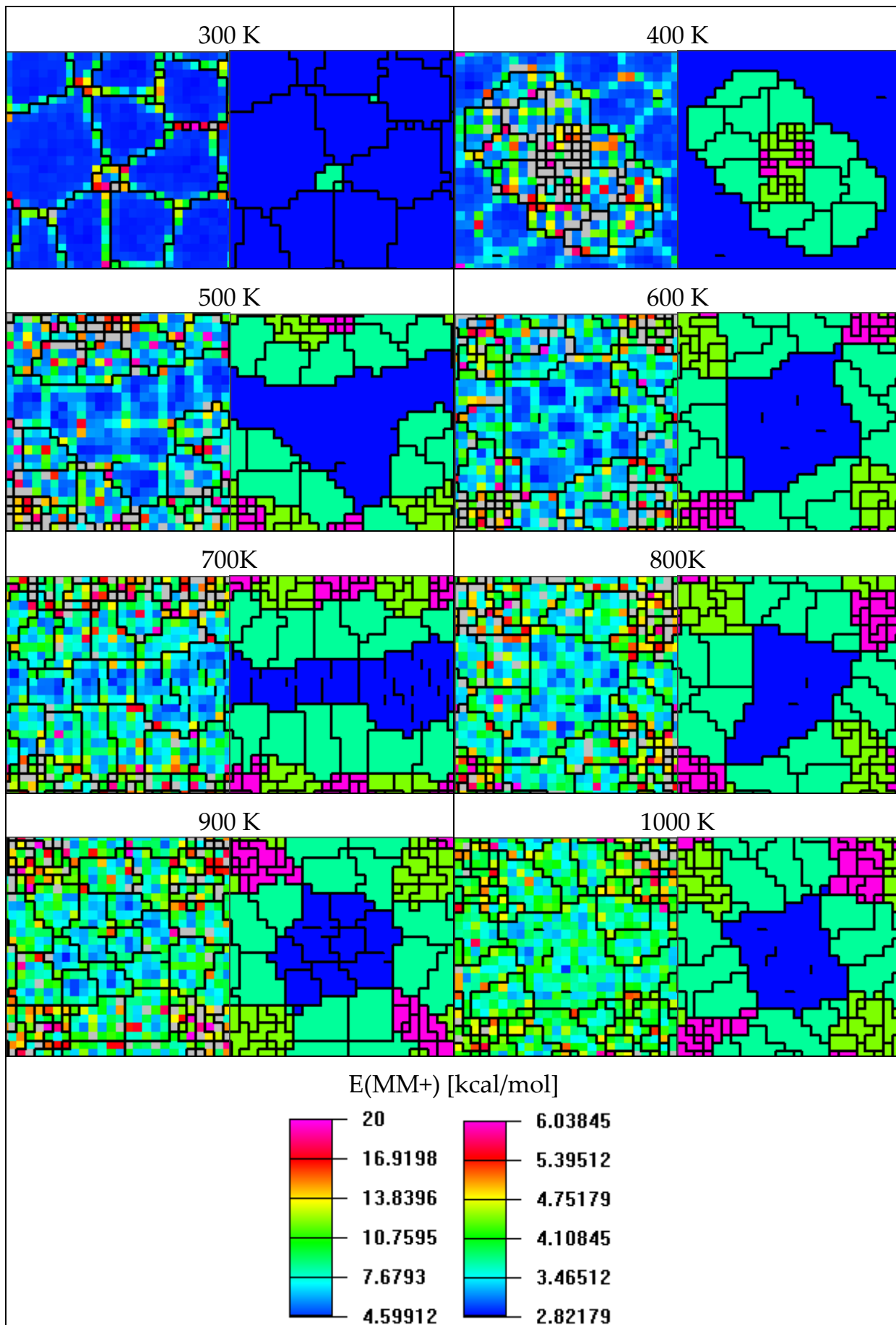


Abbildung 76: Beispielnetzarten Gridhyperfläche – Variation der Temperatur – Pentan MD, Training der kompletten Distanzmatrix, links partiell optimiert unter Verwendung der Netzrestraints nach COLORIX1, rechts vollständig optimiert nach COLORIX2

Offensichtlich konkurrieren auch bei höheren Temperaturen zwei verschiedene Netzkonformationen mit fast identischem RMSG-Wert. Zu erkennen sind in Abbildung 76:

- die *anti/anti* (dunkelblaue) Numerationskonformere als Band (500 K, 700 K)
- die *anti/anti* (dunkelblaue) Numerationskonformere als Agglomerat (600 K, 800 K, 900 K, 1000 K)

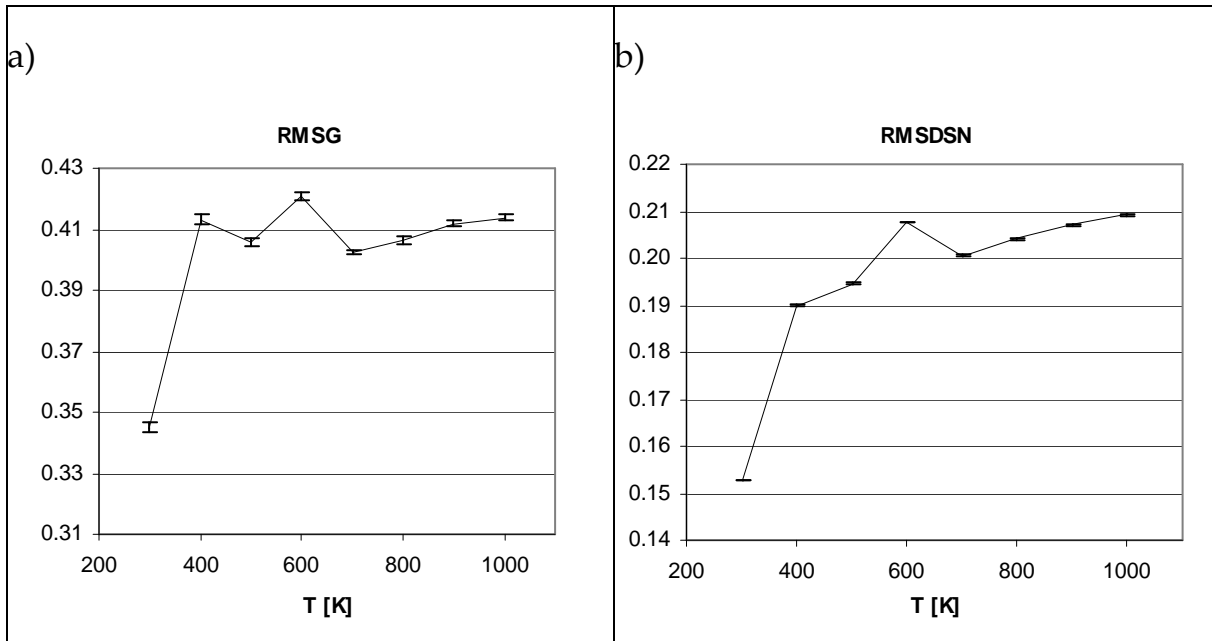
#### 2.8.2.4 MD Pentan, Training der Schweratomdistanzen

Das Training der Schweratomsmatrix ergibt wiederum eine vereinfachte Hyperfläche, da Rotationsenantiomere auf Grund von C-C Abständen im Molekül nicht unterscheidbar sind. Wiederum lässt sich der RMSDCF-Wert nicht sinnvoll ermitteln. Die statistischen Daten des Netztrainings der Moleküldynamiken aus Kapitel 2.8.2.2 sind in Tabelle 32 aufgeführt. [25x25 Neuronen; 100 *epochs*;  $\Delta_{\max}(1)=12$ ;  $f_0=0.4$ ; Abstandsfunktion parabolisch, Training der C-C Abstände]

**Tabelle 32:** Pentan MD-Simulationen bei unterschiedlichen Temperaturen, Training der Schweratommatrix

T [K]	Ø RMSG	$\sigma$	relerr%	min	max	ndata
1000	0.41375	0.00098	0.24	0.41172	0.41533	10
900	0.41171	0.00094	0.23	0.41023	0.41315	10
800	0.40622	0.00140	0.34	0.40362	0.40781	10
700	0.40258	0.00076	0.19	0.40135	0.40392	10
600	0.42094	0.00118	0.28	0.41956	0.42306	10
500	0.40556	0.00144	0.36	0.40263	0.40719	10
400	0.41327	0.00134	0.32	0.41057	0.41563	10
300	0.34498	0.00144	0.42	0.34255	0.34682	10
T [K]	Ø RMSDSN	$\sigma$	relerr%	min	max	ndata
1000	0.20920	0.00016	0.08	0.20889	0.20938	10
900	0.20710	0.00012	0.06	0.20693	0.20729	10
800	0.20420	0.00020	0.10	0.20389	0.20445	10
700	0.20058	0.00017	0.08	0.20037	0.20091	10
600	0.20766	0.00017	0.08	0.20743	0.20792	10
500	0.19460	0.00014	0.07	0.19438	0.19481	10
400	0.18994	0.00013	0.07	0.18975	0.19018	10
300	0.15277	0.00014	0.09	0.15247	0.15302	10

Die grafische Auftragung des RMSG-Wertes in Abbildung 77 zeigt wiederum den typischen Knick, diesmal bei 400 K, ab dem alle Konformationen im Trainingssatz enthalten sind. Wahrscheinlich liegt diese Temperatur durch die mangelnde Unterscheidbarkeit der Rotationsisomere niedriger als in obigen Trainingseinheiten mit der vollen Distanzmatrix und den Diederwinkeln.



**Abbildung 77: Variation der Temperatur – Pentan MD, Training der Schweratomdistanzmatrix**

Beispiele der erhaltenen Netzarten sind in Abbildung 78 angeführt. Die Nichtunterscheidbarkeit energetisch unterschiedlicher Konformationen, z.B.  $g^+g^-$  und  $g^-g^+$ , siehe Abbildung 78 bei 1000 K, führt hier zum Effekt, dass diese Konformationen stochastisch über eine gemeinsame Fläche verteilt sind. Diese gemeinsame Fläche ist hier farblich grün und magentafarben gesprenkelt. Außerdem fällt die größere „Unruhe“ im erhaltenen Bild auf.

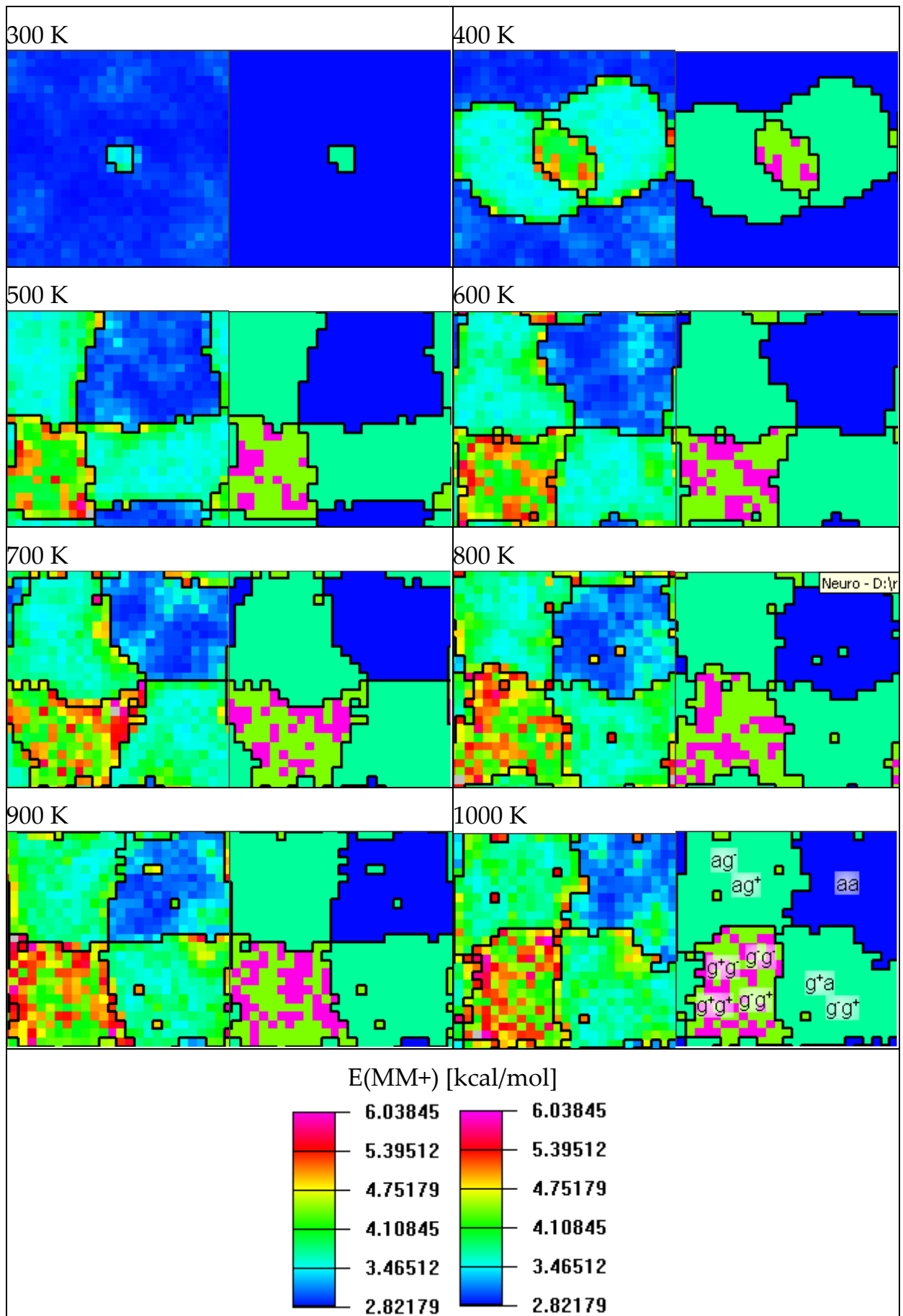
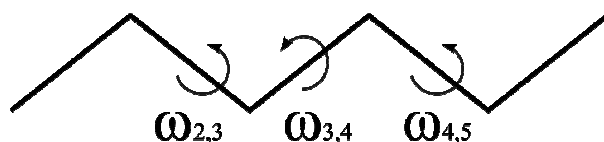
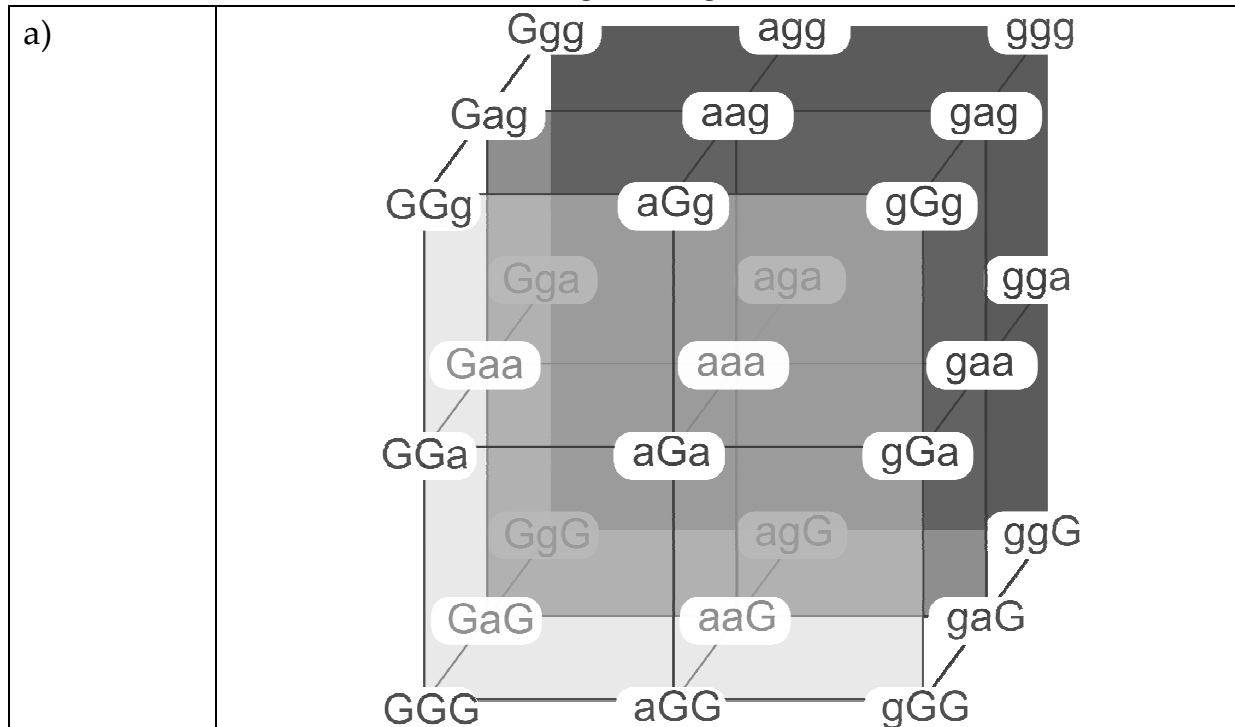
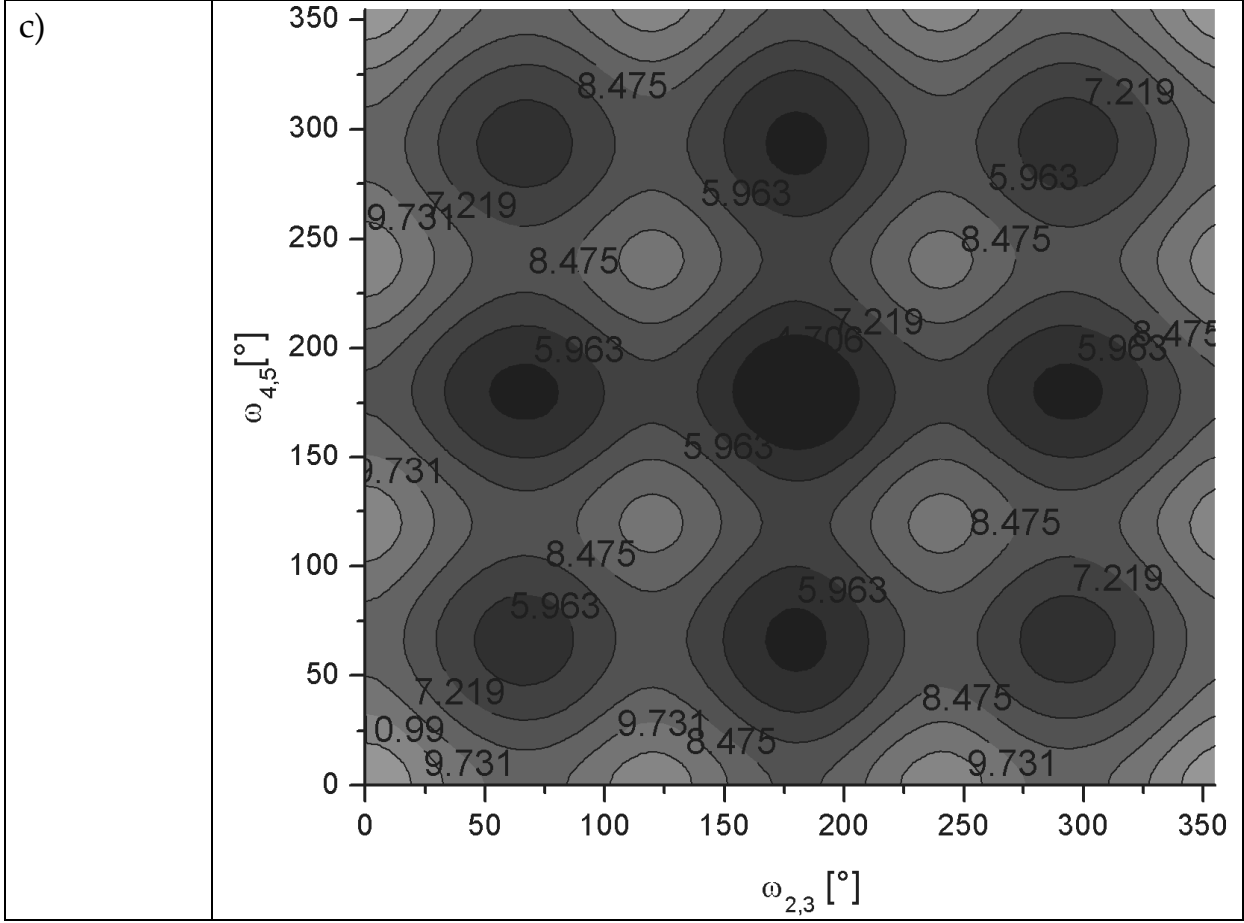
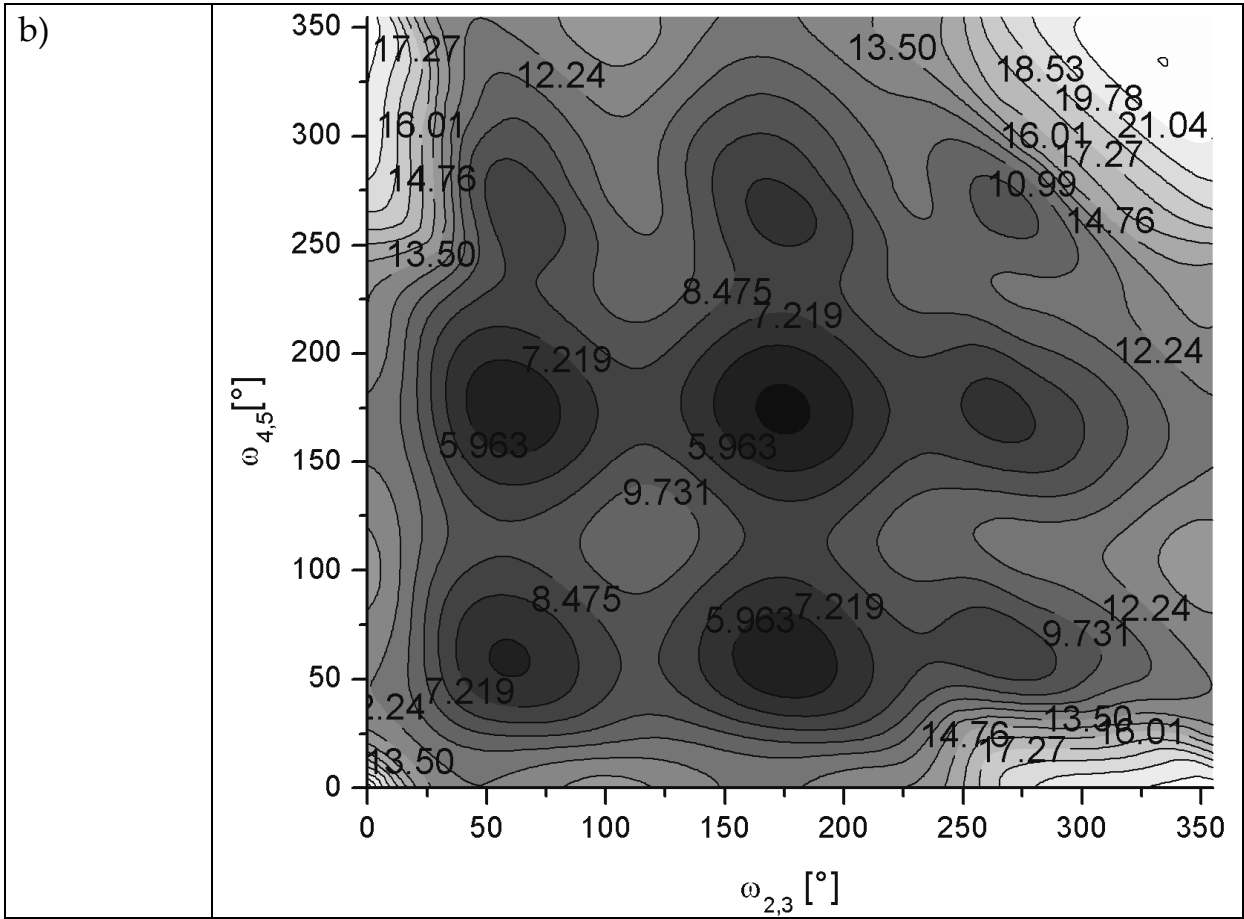


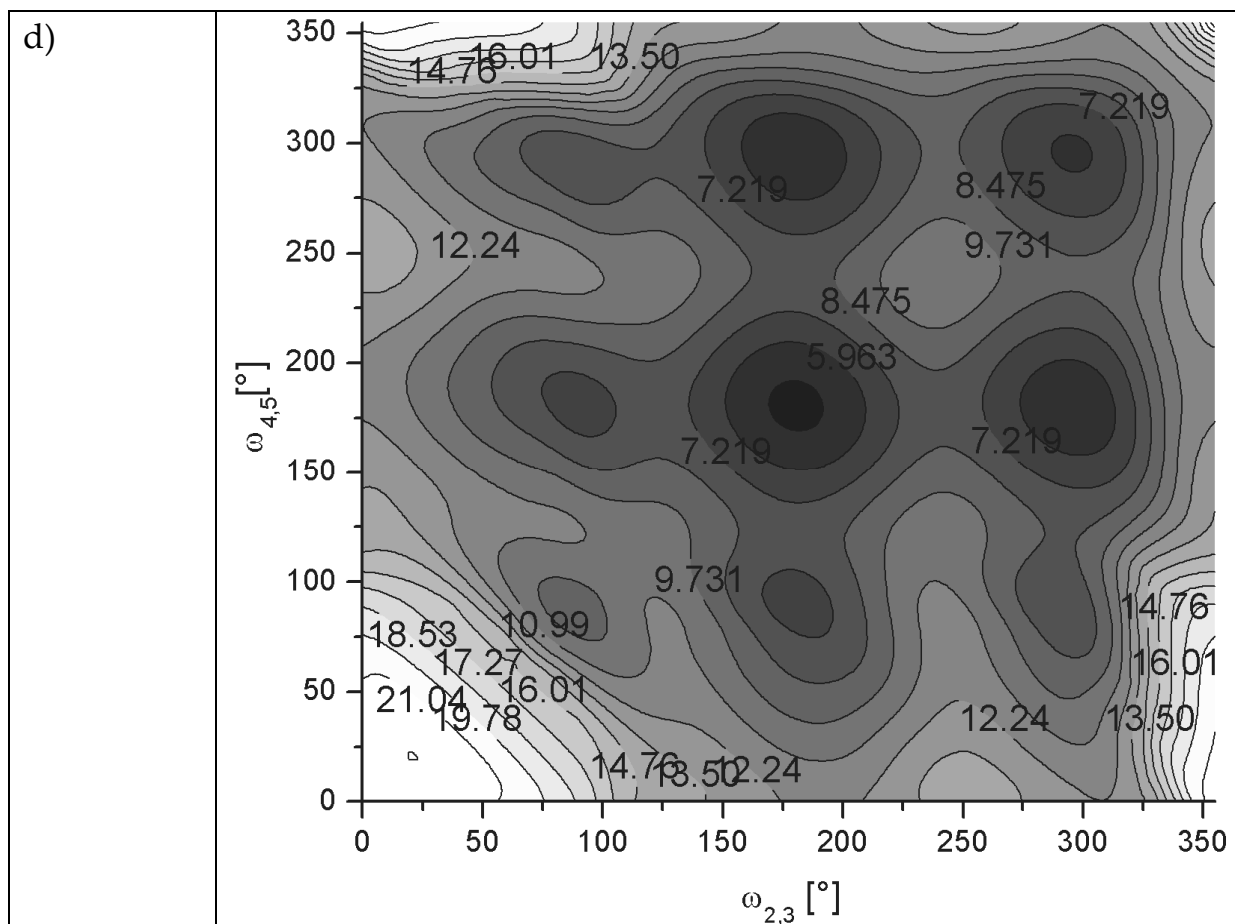
Abbildung 78: Beispielnetzkarten Gridhyperfläche – Variation der Temperatur – Pentan MD, Training der Schweratomdistanzmatrix, links partiell optimiert unter Verwendung der Netzrestraints nach COLORIX1, rechts vollständig optimiert nach COLORIX2

2.8.3 *n*-HexanAbbildung 79: CC-Diederwinkel im *n*-Hexan

Das Hexanmolekül enthält ohne die Berücksichtigung der freien Drehbarkeit der terminalen Methylgruppen bereits 3 frei drehbare Diederwinkel (siehe Abbildung 79) und damit eine Gesamtzahl von  $3^3=27$  Konformationen. Die Anordnung dieser Minima auf der Hyperfläche ist dreidimensional, eine Darstellung in den zweidimensionalen Raum ist auf klassische Weise nicht ohne weiteres möglich. Es liegt mithin eine echte Anwendung für die Projektion aus höherer Dimension vor. Wenn der zentrale CC-Torsionswinkel des Hexans ( $\omega_{3,4}$ ) in den drei ungefähren Minimumspositionen  $60^\circ$  ( $G = gauche^+$ ),  $180^\circ$  ( $a = anti$ ) und  $-60^\circ$  ( $g = gauche^-$ ), gehalten wird, kann die dreidimensionale Hyperfläche auf drei zweidimensionale Schnitte reduziert werden. Die mittels des Kraftfeldes MM+ des Programms HYPERCHEM sowie des Skriptes GRID.TCL (siehe Kapitel 5.2.2) erhaltenen Energiekonturdiagramme, durch die Variation der äußeren CC-Torsionswinkel ( $\omega_{2,3}$ ,  $\omega_{4,5}$ ) sowie ihre dreidimensionale Konnektivität sind in Abbildung 80 dargestellt.







**Abbildung 80:** a) Schematische Darstellung der drei Schnitte der Hexan-Hyperfläche, Isolinien der Energie (Kraftfeld MM+) in kcal/mol  
 b) MM+ Isooberflächenschnitt für  $\omega_{3,4}=60^\circ$  (G)  
 c) MM+ Isooberflächenschnitt für  $\omega_{3,4}=180^\circ$  (a)  
 d) MM+ Isooberflächenschnitt für  $\omega_{3,4}=-60^\circ$  (g)

Mit Hilfe des Programms PRINT\_SELECTION\_VALUES\_GUI.TK werden exemplarisch die Konformationen zu einer Netzkarte, erzeugt aus MD-Daten einer Simulation bei 500 K, zugeordnet, siehe Abbildung 82. Der weitgehende Erhalt der Nachbarschaftsbeziehungen ist gut zu erkennen. Interessanterweise wird die *all-anti* Konformation zweimal auf der Hyperfläche abgebildet. Außerdem spalten sich einige Konformationen (gGa, aGg, Gga, agG) aufgrund sterischer Wechselwirkungen in zwei energiegeliche Minima, andere (Ggg, GGg, gGG, ggG) spalten sich ebenfalls aufgrund sterischer Wechselwirkungen in energetisch unterschiedliche Minima.



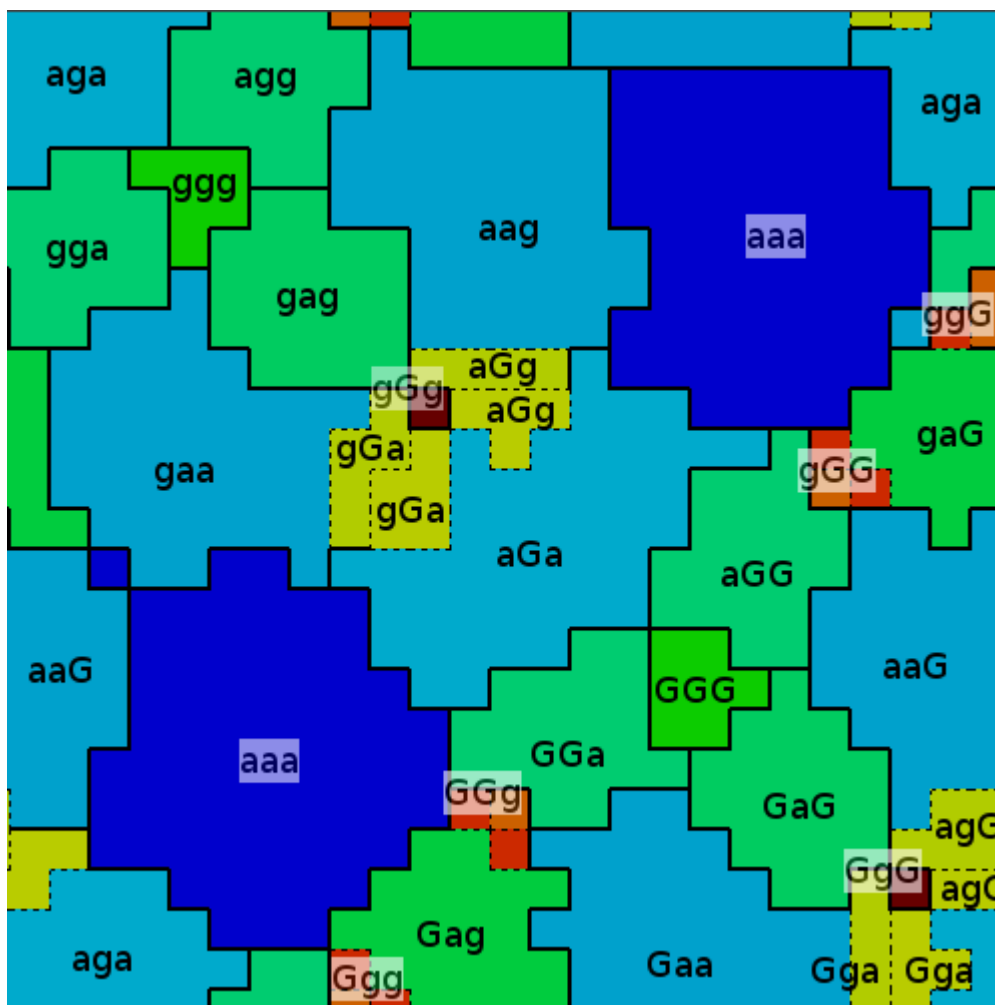


Abbildung 81: Beispielprojektion der MM+ Hyperfläche des Hexans durch Training der CC-Diederwinkel samt Energieskala in kcal/mol

### 2.8.3.1 Training der Grid-Hyperfläche

Es wird zunächst ein Referenzdatensatz erstellt, welcher die komplette durch Rotation der CC-Diederwinkel erreichbare Hyperfläche beinhaltet.

Variiert werden in einer Gridrechnung mit dem Programm GRID.TCL (5.2.2) unter Verwendung des Kraftfelds MM+ (HYPERCHEM) folgende CC-Diederwinkel

- $\omega_{2,3}$ : 0, 15, 30, ..., 345
- $\omega_{3,4}$ : 0, 15, 30, ..., 345
- $\omega_{4,5}$ : 0, 15, 30, ..., 345

Die einzelnen Strukturen werden wie bereits beim Butan und Pentan optimiert, wobei besagte Diederwinkel während der Geometrieoptimierung eingefroren bleiben. Damit sind auch obige Schnitte der dreidimensionalen Hyperfläche aus Abbildung 80 im Referenzsatz von insgesamt 13825 Strukturen enthalten.

Für die Ermittlung der RMSDCF-Werte wird eine Klassifizierung der Hexan-konformationen mittels des Programms CLASSIFY\_SNP\_CONFORMATIONS.EXE, siehe Kapitel 5.3.5, durchgeführt.

Trainiert werden die Kohonennetze in den folgenden Unterkapiteln mit Vorgabewerten von  $f_0=0.15$ ,  $epochs=1000$ ,  $neighborhood\ function =$  parabolisch,  $NETX=NETY=21$ ,  $linear\ learning\ decay$  für die nichtvariieren Netzparameter.

### 2.8.3.1.1 Variation der epochs

Tabelle 33: Hexan - Variation der epochs

epochs	Ø RMSDCF	$\sigma$	relerr%	min	max	ndata
1000	0.00461	0.00077	16.69	0.00367	0.00573	5
500	0.00419	0.00067	15.94	0.00353	0.00504	5
300	0.00432	0.00062	14.30	0.00349	0.00489	5
200	0.00482	0.00078	16.19	0.00382	0.00590	5
100	0.00475	0.00046	9.67	0.00402	0.00520	5
50	0.00580	0.00069	11.84	0.00503	0.00689	5
10	0.00697	0.00115	16.52	0.00541	0.00818	5
2	0.01195	0.00135	11.33	0.00985	0.01344	5
1	0.01445	0.00156	10.79	0.01259	0.01588	5
epochs	Ø RMSG	$\sigma$	relerr%	min	max	ndata
1000	0.24387	0.00071	0.29	0.24274	0.24440	5
500	0.24343	0.00075	0.31	0.24266	0.24426	5
300	0.24128	0.00139	0.58	0.23971	0.24351	5
200	0.23899	0.00138	0.58	0.23774	0.24058	5
100	0.22943	0.00081	0.35	0.22848	0.23072	5
50	0.21752	0.00177	0.81	0.21581	0.22039	5
10	0.19756	0.00092	0.47	0.19612	0.19868	5
2	0.18288	0.00622	3.40	0.17515	0.19168	5
1	0.17927	0.00968	5.40	0.16663	0.18800	5
epochs	Ø RMSDSN	$\sigma$	relerr%	min	max	ndata
1000	0.06488	0.00010	0.16	0.06479	0.06502	5
500	0.06525	0.00010	0.15	0.06514	0.06538	5
300	0.06553	0.00007	0.10	0.06542	0.06559	5
200	0.06600	0.00016	0.25	0.06581	0.06624	5
100	0.06777	0.00008	0.12	0.06770	0.06788	5
50	0.07125	0.00016	0.22	0.07101	0.07145	5
10	0.08172	0.00024	0.29	0.08134	0.08194	5
2	0.09563	0.00085	0.89	0.09473	0.09701	5
1	0.10302	0.00311	3.02	0.10045	0.10810	5

Es wird wie zuvor beim Pentan eine grafische Auftragung der Qualitätsparameter durchgeführt, siehe Abbildung 82. Ein signifikanter Unterschied liegt im unterschiedlichen qualitativen Verlauf der RMSG-Kurve. Diese steigt im Gegensatz zur RMSDSN-Kurve im Verlauf des Trainingsvorgangs an, was einer erhofften Verbesserung des Qualitätsparameters zuwiderläuft. Erklärt werden kann dies damit, dass zum ersten Mal der Trainingssatz signifikant

umfangreicher ist als die zur Projektion vorhandenen 441 Neuronen. Wie bereits bei Einführung der Qualitätsparameter angemerkt, stehen RMSG- und RMSDSN-Wert hier in Konkurrenz. Anscheinend wird die Hyperfläche des Trainingsatzes im Verlaufe des Trainings gleichmäßiger durch die Projektion abgedeckt, was sich in verbesserten RMSDSN-Werten niederschlägt. Der  $\text{RMSG}_{\text{end}} = 0.24396 \pm 0.00027$  wird per Extrapolation<sup>[104]</sup> mit dem Programm Origin7.5<sup>[105]</sup> für  $\text{epochs} > 10$  ermittelt.  $\text{Epoch}_{\text{end}}(\text{RMSG})$  ist mit maximal 1% Abweichung bei 500 *epochs* erreicht.  $\text{RMSDSN}_{\text{end}} = 0.06482 \pm 0.00007$  wurde ebenfalls per Extrapolation<sup>[106]</sup> bestimmt.  $\text{Epoch}_{\text{end}}(\text{RMSDSN})$  ist mit maximal 1% Abweichung ebenfalls bei 500 *epochs* erreicht

$$\text{MTS}(\text{RMSG}, 1\%) = \text{MTS}(\text{RMSDSN}, 1\%) = 13825 * 500 = 6912000$$

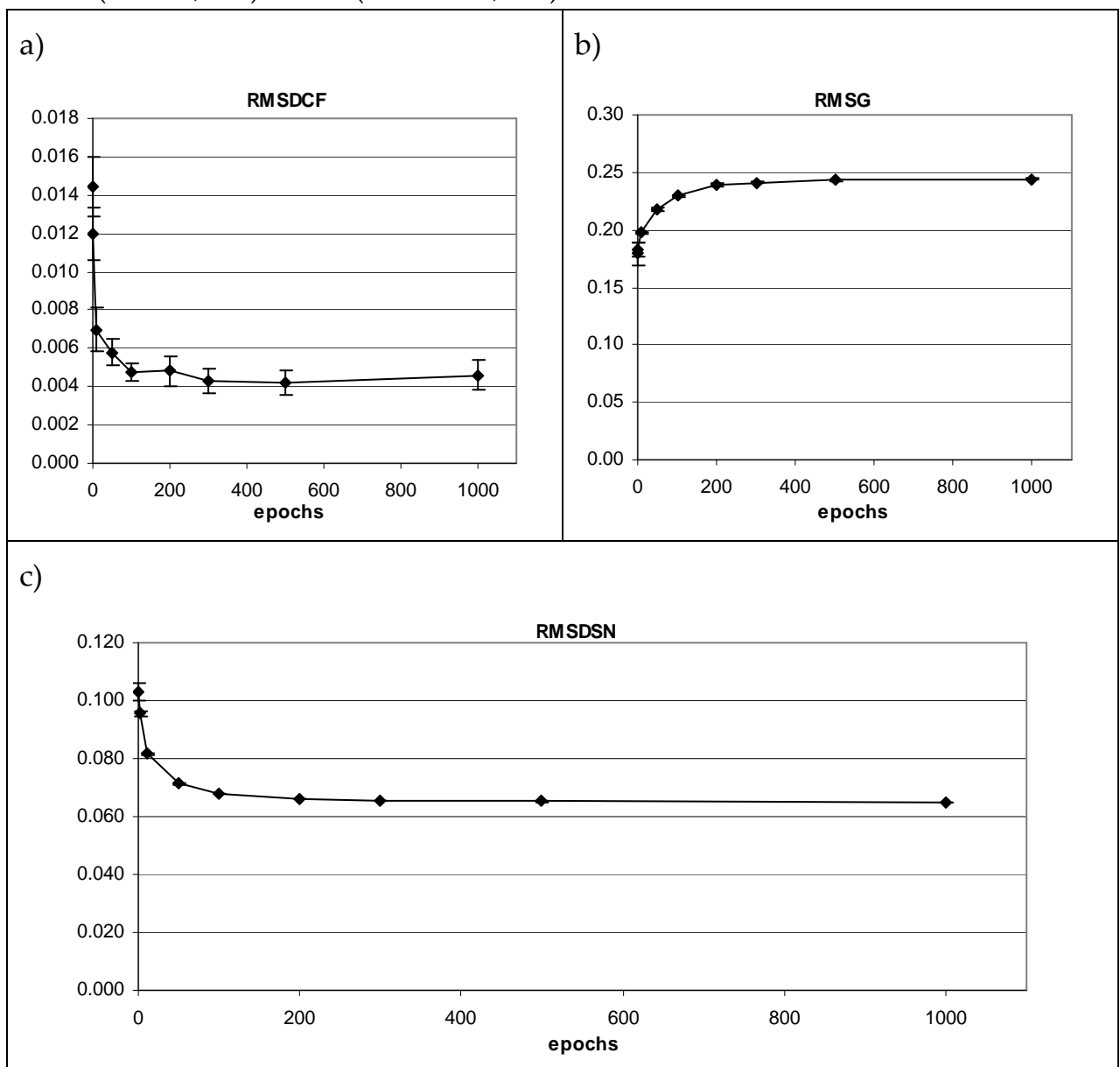


Abbildung 82: Variation der *epochs*, a) RMSDCF, b) RMSG, c) RMSDSN

2.8.3.1.2 Variation von  $f_0$ Tabelle 34: Variation des maximalen Adaptionfaktors  $f_0$  nach Gleichung (2.12)

$f_0$	Ø RMSDCF	$\sigma$	relerr%	min	max	ndata
0.05	0.00527	0.00074	13.96	0.00449	0.00611	5
0.1	0.00427	0.00037	8.61	0.00400	0.00489	5
0.15	0.00477	0.00059	12.30	0.00408	0.00558	5
0.2	0.00483	0.00022	4.50	0.00450	0.00510	5
0.25	0.00550	0.00100	18.24	0.00448	0.00691	5
0.3	0.00485	0.00076	15.59	0.00375	0.00568	5
0.4	0.00468	0.00071	15.23	0.00364	0.00541	5
0.5	0.00466	0.00030	6.44	0.00434	0.00502	5
0.6	0.00407	0.00028	6.92	0.00373	0.00447	5
0.7	0.00436	0.00092	21.05	0.00339	0.00533	5
0.8	0.00509	0.00093	18.19	0.00407	0.00615	5
0.9	0.00513	0.00066	12.80	0.00409	0.00569	5
$f_0$	Ø RMSG	$\sigma$	relerr%	min	max	ndata
0.05	0.24159	0.00197	0.82	0.23886	0.24337	5
0.1	0.24318	0.00127	0.52	0.24211	0.24533	5
0.15	0.24478	0.00032	0.13	0.24444	0.24522	5
0.2	0.24550	0.00045	0.18	0.24488	0.24587	5
0.25	0.24583	0.00080	0.33	0.24459	0.24671	5
0.3	0.24570	0.00080	0.33	0.24469	0.24654	5
0.4	0.24632	0.00153	0.62	0.24422	0.24821	5
0.5	0.24878	0.00282	1.13	0.24543	0.25177	5
0.6	0.24819	0.00322	1.30	0.24369	0.25238	5
0.7	0.24915	0.00436	1.75	0.24437	0.25400	5
0.8	0.24715	0.00236	0.95	0.24523	0.25107	5
0.9	0.24818	0.00274	1.10	0.24484	0.25231	5
$f_0$	Ø RMSDSN	$\sigma$	relerr%	min	max	ndata
0.05	0.06559	0.00007	0.11	0.06547	0.06566	5
0.1	0.06513	0.00011	0.17	0.06503	0.06526	5
0.15	0.06497	0.00001	0.02	0.06496	0.06499	5
0.2	0.06477	0.00006	0.10	0.06470	0.06485	5
0.25	0.06468	0.00011	0.17	0.06449	0.06478	5
0.3	0.06463	0.00004	0.06	0.06457	0.06467	5
0.4	0.06448	0.00004	0.06	0.06444	0.06454	5
0.5	0.06435	0.00006	0.10	0.06429	0.06443	5
0.6	0.06428	0.00006	0.10	0.06419	0.06437	5
0.7	0.06417	0.00004	0.07	0.06411	0.06422	5
0.8	0.06414	0.00005	0.08	0.06406	0.06420	5
0.9	0.06406	0.00006	0.09	0.06402	0.06416	5

Eine grafische Auftragung der Tabelle 34 wird in Abbildung 83 vorgenommen. Zunächst ist darauf hinzuweisen, dass die absoluten Unterschiede geringer sind, als sie durch die Skalierung der y-Achse für die RMSG- und RMSDSN-Werte suggeriert werden. Auch hier sind RMSG-Werte und RMSDSN-Werte wieder gegenläufig, wobei die großen  $f_0$ -Werte eher zu niedrigen RMSDSN-Werten und damit zu einer kompletteren Abdeckung der

Hyperfläche führen. Das interessanteste Ergebnis liegt aber wahrscheinlich in der Standardabweichung des RMSG-Wertes. Hier gibt es ein klares Minimum bei  $f_0=0.15$ . Dies bedeutet, dass das Netztraining für diesen Parameter am ehesten reproduzierbare Werte hervorbringt. Der Graph des RMSDCF-Wertes ergibt keinen interpretationsfähigen Trend außerhalb der Fehlergrenzen.

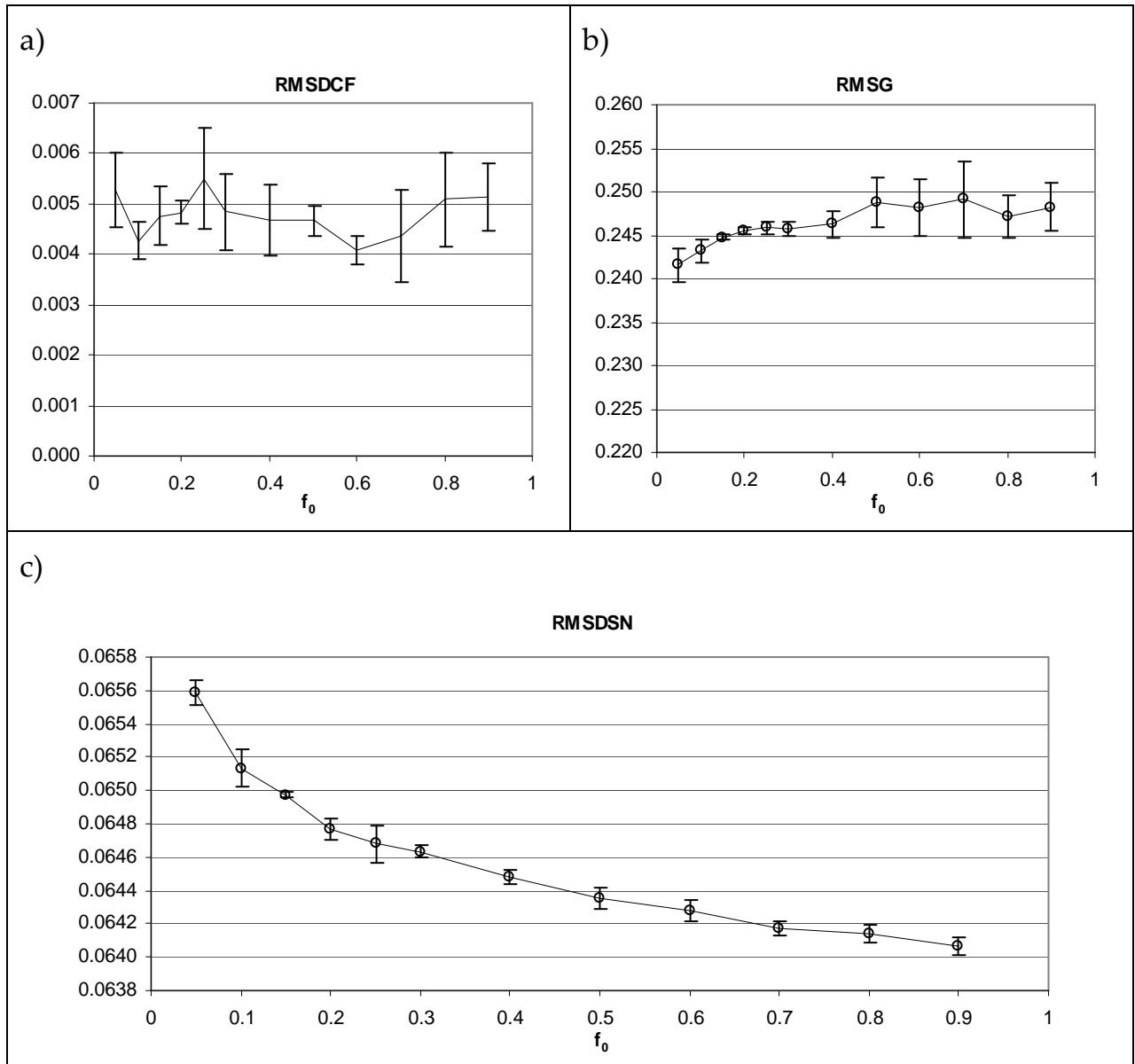


Abbildung 83: Variation von  $f_0$ , a) RMSDCF, b) RMSG, c) RMSDSN

### 2.8.3.1.3 Variation von $\Delta_{\max}(1)$

Tabelle 35: Variation von  $\Delta_{\max}(1)$  nach Gleichung (2.15)

$\Delta_{\max}(1)$	$\emptyset$ RMSDCF	$\sigma$	relerr%	min	max	ndata
10	0.00451	0.00063	13.97	0.00387	0.00556	5
9	0.00482	0.00086	17.78	0.00380	0.00617	5
8	0.00437	0.00039	8.84	0.00395	0.00493	5
7	0.00512	0.00052	10.21	0.00450	0.00572	5
6	0.00438	0.00046	10.47	0.00361	0.00481	5
5	0.00452	0.00055	12.09	0.00367	0.00500	5
4	0.00472	0.00069	14.66	0.00418	0.00562	5
3	0.00450	0.00101	22.45	0.00309	0.00582	5
2	0.00507	0.00049	9.66	0.00461	0.00585	5
1	0.00500	0.00112	22.48	0.00325	0.00597	5
$\Delta_{\max}(1)$	$\emptyset$ RMSG	$\sigma$	relerr%	min	max	ndata
10	0.24367	0.00069	0.28	0.24281	0.24440	5
9	0.24416	0.00082	0.33	0.24297	0.24497	5
8	0.24533	0.00042	0.17	0.24475	0.24585	5
7	0.24553	0.00074	0.30	0.24455	0.24623	5
6	0.24571	0.00085	0.34	0.24462	0.24669	5
5	0.24601	0.00098	0.40	0.24483	0.24678	5
4	0.24659	0.00290	1.17	0.24399	0.25157	5
3	0.24525	0.00182	0.74	0.24221	0.24668	5
2	0.25313	0.01161	4.59	0.23507	0.26604	5
1	0.32858	0.00458	1.40	0.32173	0.33462	5
$\Delta_{\max}(1)$	$\emptyset$ RMSDSN	$\sigma$	relerr%	min	max	ndata
10	0.06504	0.00004	0.06	0.06499	0.06510	5
9	0.06486	0.00009	0.13	0.06480	0.06501	5
8	0.06476	0.00007	0.11	0.06471	0.06488	5
7	0.06474	0.00006	0.10	0.06468	0.06484	5
6	0.06461	0.00009	0.14	0.06448	0.06474	5
5	0.06445	0.00005	0.07	0.06440	0.06451	5
4	0.06434	0.00005	0.08	0.06429	0.06441	5
3	0.06412	0.00009	0.15	0.06401	0.06425	5
2	0.06394	0.00009	0.14	0.06387	0.06408	5
1	0.06382	0.00010	0.15	0.06372	0.06393	5

Die grafische Auftragung der Qualitätsparameter nach Variation der Größe der anfänglich trainierten Nachbarschaft  $\Delta_{\max}(1)$  ist in Abbildung 84 vorgenommen. Die RMSDCF-Werte ergeben im Rahmen der Fehlergrenzen keinen erkennbaren Trend. Ansonsten sind die Ergebnisse konsistent zu den in Kapitel 2.8.1.5.3 und Kapitel 2.8.2.1.3 erhaltenen. Der RMSG-Graph ergibt ähnlich wie beim Butan eine „untere Schranke“ für bei  $\Delta_{\max}(1) = 3$  (30%). Die Trends der RMSG- und RMSDSN-Werte sind wiederum gegenläufig, was durch die Skalierung der Achsen im RMSG-Graphen nicht zu sehen ist, aber an den Tabellenwerten abgelesen werden kann. Augenscheinlich bewirkt eine kleinere Nachbarschaft einen besseren Erhalt der Diversität der repräsentierten Strukturen aus der ursprünglichen zufällig erstellten Anfangskonfigurati-

on des Netzes. Ab  $\Delta_{\max}(1) = 5$  (50% des Maximalwertes) liegt der relative Fehler der RMSG-Werte unter 0.5%, was als zweite untere Grenze angesehen werden kann, um reproduzierbar ähnliche Strukturen zu erhalten.

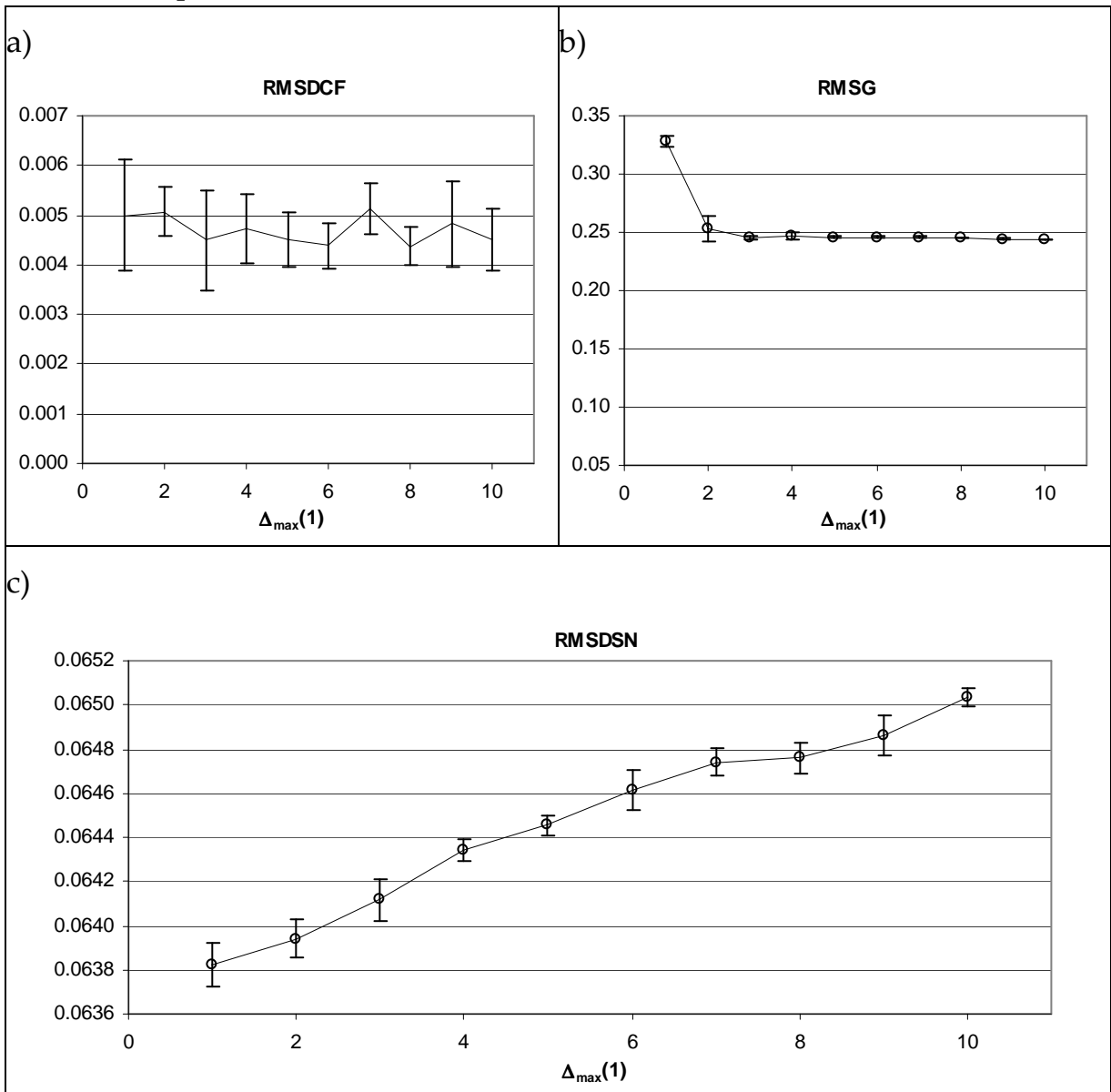


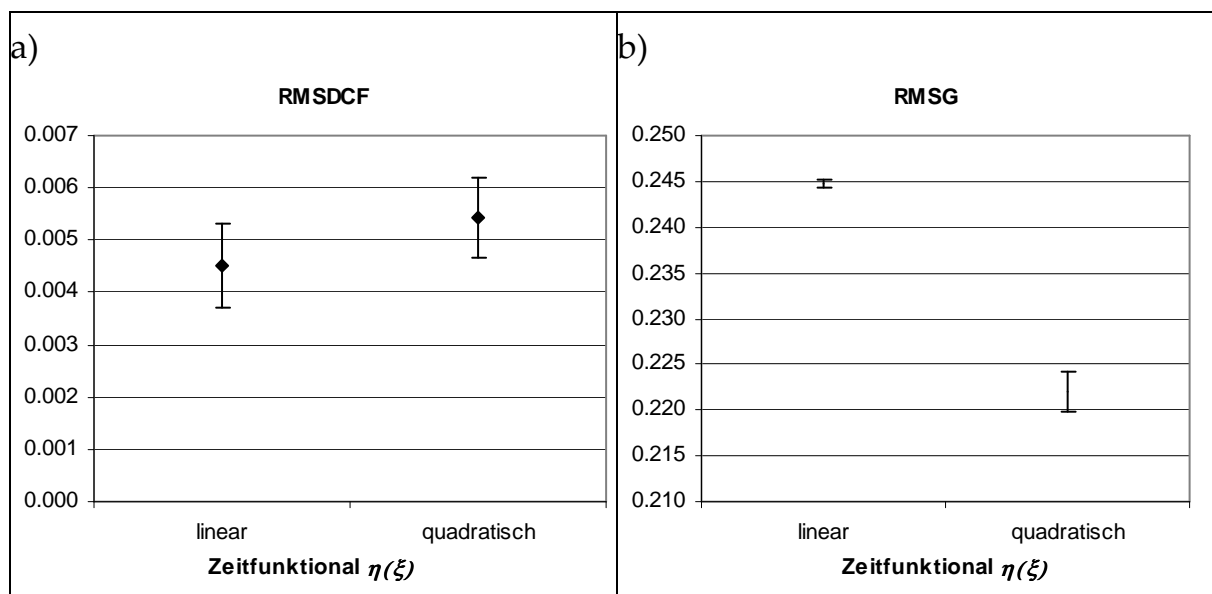
Abbildung 84: Variation von  $\Delta_{\max}(1)$ , a) RMSDCF, b) RMSG, c) RMSDSN

### 2.8.3.1.4 Variation von $\eta(\xi)$

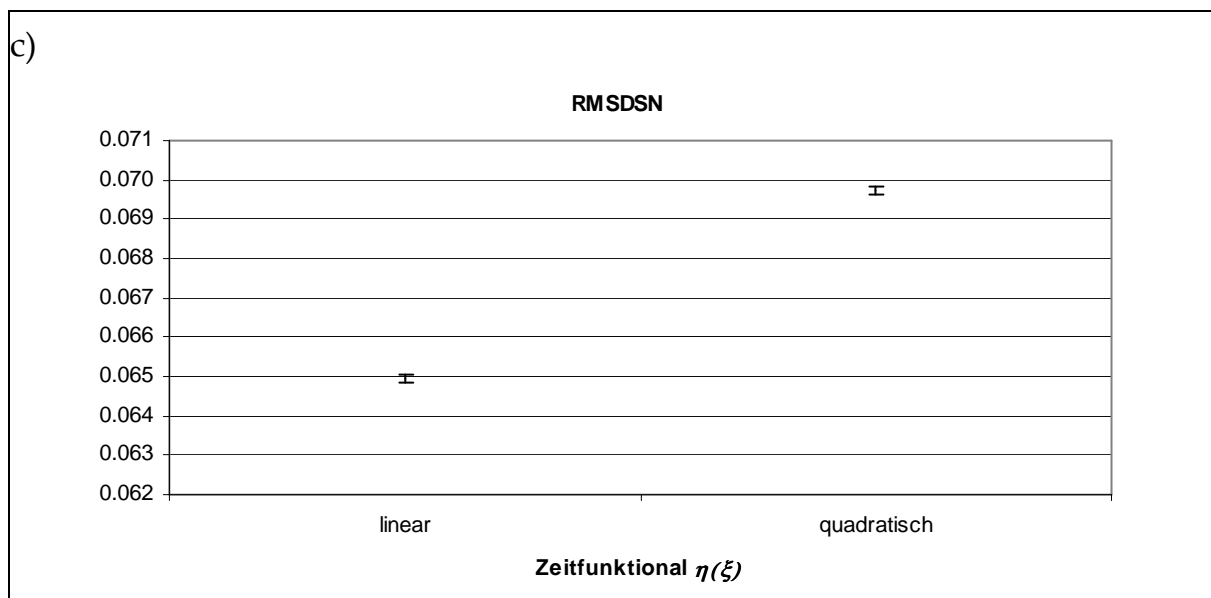
**Tabelle 36:** Variation der Zeitabhängigkeit des Adaptionfaktors,  $\eta(\xi)$  nach Gleichung (2.17)

$\eta(\xi)$ (2.17)	$\emptyset$ RMSDCF	$\sigma$	relerr%	min	max	ndata
a) linear	0.00453	0.00080	17.7	0.00351	0.00564	5
b) quadratisch	0.00542	0.00077	14.2	0.00430	0.00614	5
$\eta(\xi)$ (2.17)	$\emptyset$ RMSG	$\sigma$	relerr%	min	max	ndata
a) linear	0.24472	0.000467	0.2	0.24422	0.24547	5
b) quadratisch	0.22208	0.002211	1.0	0.21959	0.22401	5
$\eta(\xi)$ (2.17)	$\emptyset$ RMSDSN	$\sigma$	relerr%	min	max	ndata
a) linear	0.06496	0.000095	0.1	0.06483	0.06507	5
b) quadratisch	0.06972	0.000088	0.1	0.06957	0.06979	5

Der grafische Auftragung obiger Tabelle in Abbildung 85 zeigt aufgrund kleinerer Fehler einen deutlicheren Effekt als in den vorangehenden Kapiteln. Wie beim Butan und im Gegensatz zum Pentan sind dabei insgesamt Vorteile für die lineare Funktion erkennbar. Der RMSG-Wert liegt zwar höher, ist dafür aber mit deutlich kleinerem Fehler behaftet.





Abbildung 85: Variation von  $\eta(\xi)$  nach Gleichung (2.17)

### 2.8.3.1.5 Variation von NETX, NETY

Tabelle 37: Variation der Netzgröße NETX=NETY

NETX=NETY	$\bar{\sigma}$ RMSDCF	$\sigma$	relerr%	min	max	ndata
31	0.00334	0.00054	16.26	0.00289	0.00420	5
21	0.00434	0.00063	14.60	0.00366	0.00502	5
11	0.00953	0.00172	18.06	0.00721	0.01132	5
5	0.02783	0.00316	11.36	0.02389	0.03234	5
NETX=NETY	$\bar{\sigma}$ RMSG	$\sigma$	relerr%	min	max	ndata
31	0.18520	0.00086	0.46	0.18448	0.18645	5
21	0.24413	0.00082	0.33	0.24311	0.24528	5
11	0.38343	0.00072	0.19	0.38222	0.38403	5
5	0.64571	0.00632	0.98	0.63694	0.65225	5
NETX=NETY	$\bar{\sigma}$ RMSDSN	$\sigma$	relerr%	min	max	ndata
31	0.05016	0.00009	0.18	0.05001	0.05022	5
21	0.06496	0.00009	0.13	0.06486	0.06506	5
11	0.10018	0.00015	0.15	0.10002	0.10041	5
5	0.16767	0.00027	0.16	0.16732	0.16794	5

Die obige Tabelle ist in Abbildung 86 grafisch aufgetragen. Konsistent zu den vorhergehenden Molekülen steigt die Qualität mit größeren Netzarten asymptotisch an.

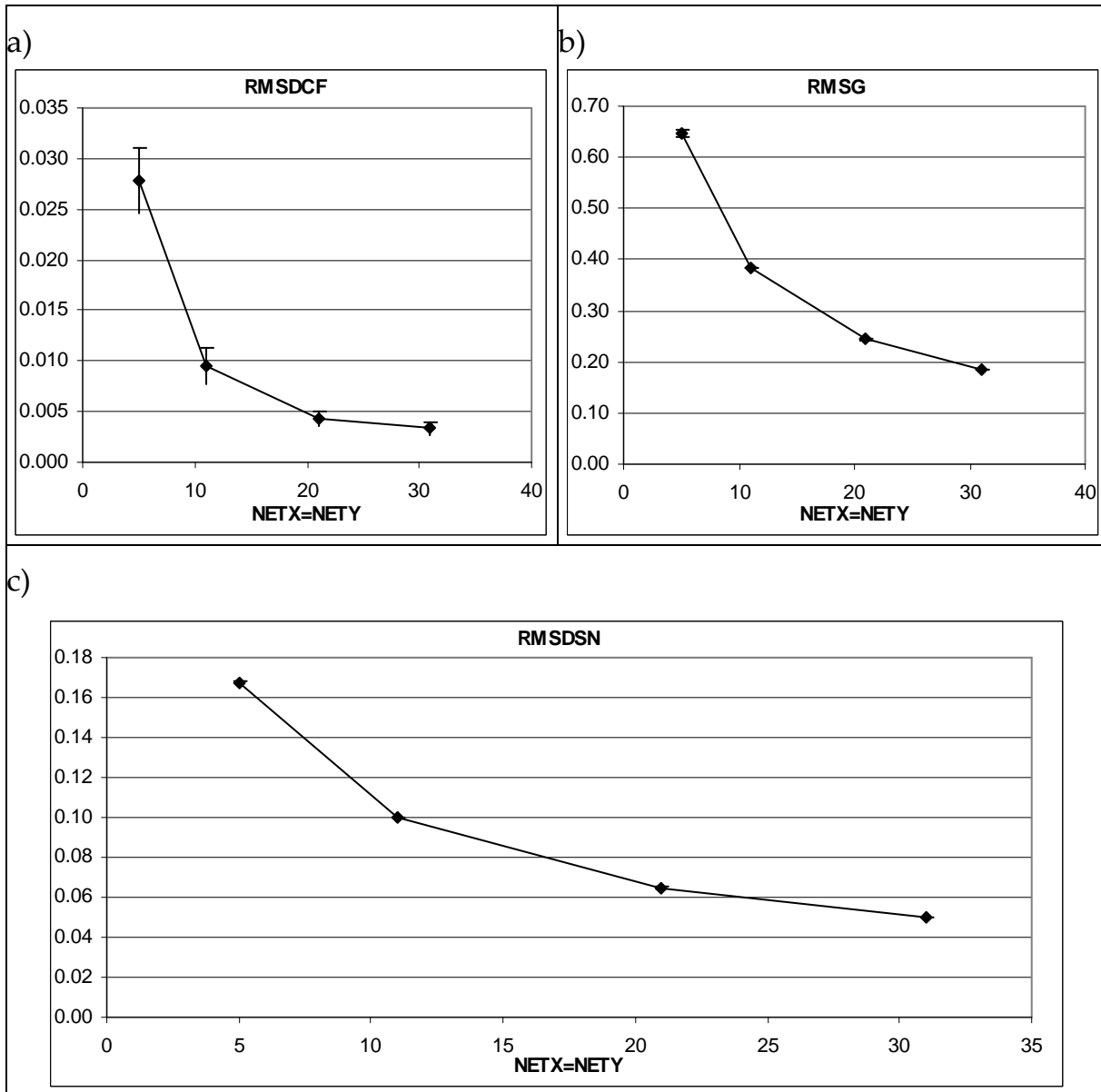


Abbildung 86: Variation der Netzgröße NETX=NETY

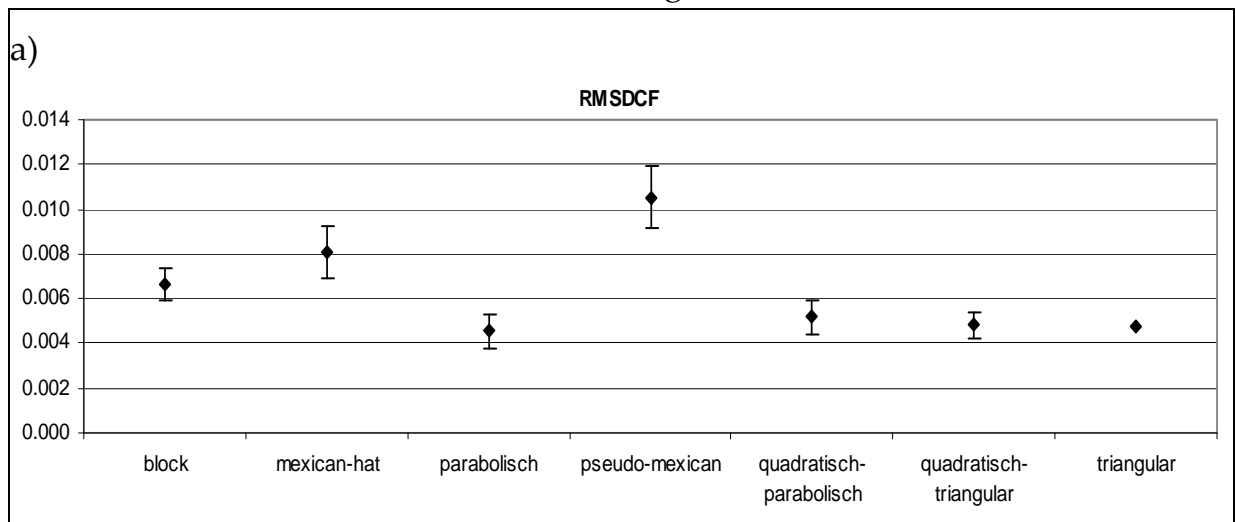
### 2.8.3.1.6 Variation der Abstandsfunktion $\vartheta(\zeta)$

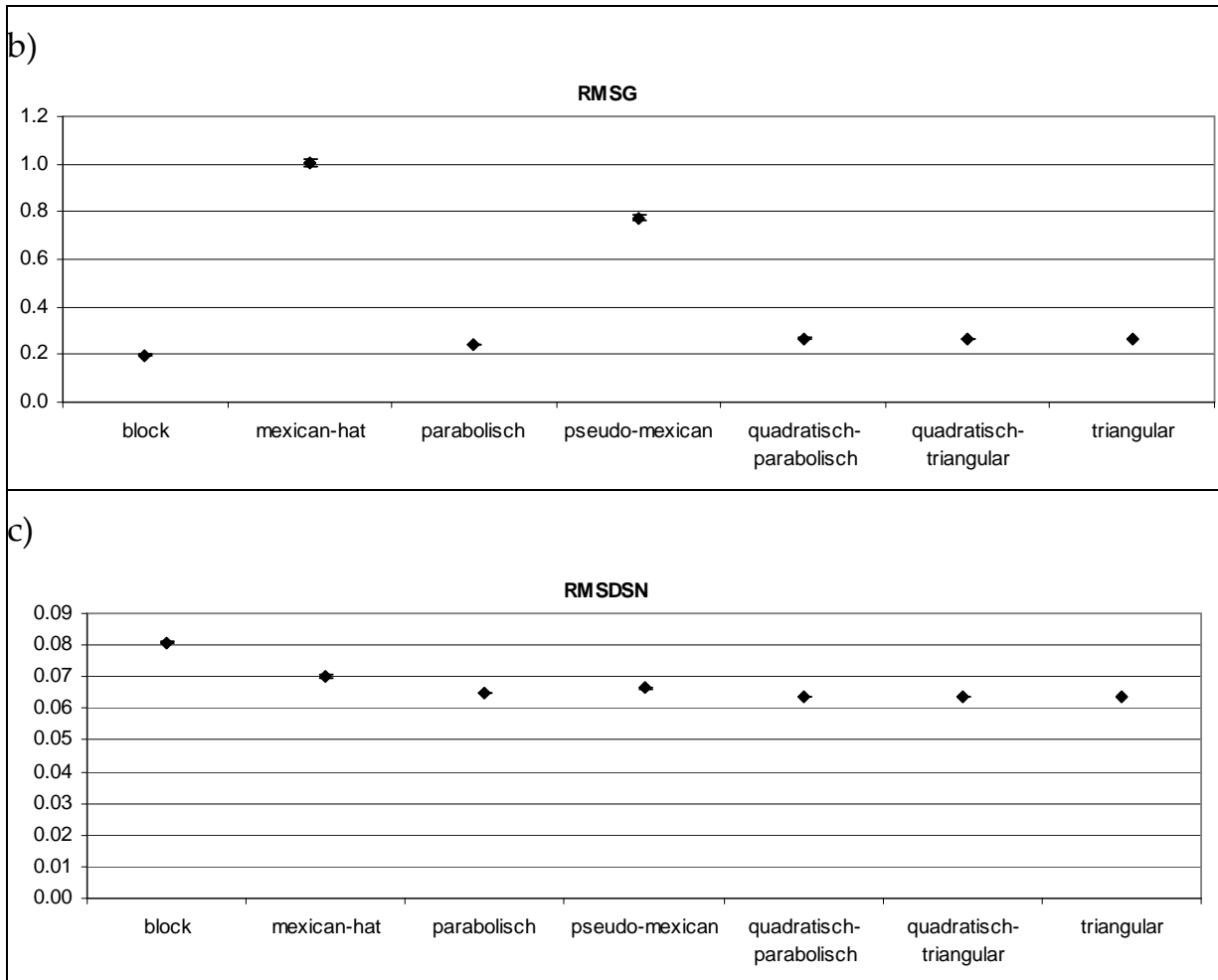
Tabelle 38: Variation der Abstandsfunktion  $\vartheta(\zeta)$  nach Gleichung (2.18)

$\vartheta(\zeta)$	$\bar{\sigma}$ RMSDCF	$\sigma$	relerr%	min	max	ndata
<i>block</i>	0.00665	0.00073	10.93	0.00607	0.00771	5
<i>Mexican hat</i>	0.00807	0.00116	14.33	0.00694	0.00969	5
<i>parabolisch</i>	0.00454	0.00075	16.53	0.00381	0.00577	5
<i>pseudo Mexican</i>	0.01054	0.00138	13.05	0.00870	0.01234	5
<i>quadratisch-parabolisch</i>	0.00517	0.00076	14.78	0.00410	0.00602	5
<i>quadratisch-triangular</i>	0.00481	0.00059	12.35	0.00426	0.00564	5
<i>triangular</i>	0.00476	0.00112	23.	0.00326	0.006169	5

$\vartheta(\zeta)$	$\varnothing$ RMSG	$\sigma$	relerr%	min	max	ndata
<i>block</i>	0.19845	0.00220	1.11	0.19456	0.19979	5
<i>Mexican hat</i>	1.00406	0.01645	1.64	0.98291	1.02607	5
parabolisch	0.24399	0.00084	0.35	0.24273	0.24487	5
<i>pseudo Mexican</i>	0.77411	0.01028	1.33	0.76336	0.78680	5
quadratisch- parabolisch	0.26798	0.00106	0.40	0.26664	0.26895	5
quadratisch- triangular	0.26733	0.00076	0.29	0.26660	0.26830	5
triangular	0.26825	0.000974039	0.36	0.26716	0.26960	5
$\vartheta(\zeta)$	$\varnothing$ RMSDSN	$\sigma$	relerr%	min	max	ndata
<i>block</i>	0.08082	0.00013	0.16	0.08065	0.08096	5
<i>Mexican hat</i>	0.06997	0.00046	0.66	0.06957	0.07048	5
parabolisch	0.06498	0.00008	0.13	0.06489	0.06508	5
<i>pseudo Mexican</i>	0.06653	0.00022	0.33	0.06622	0.06674	5
quadratisch- parabolisch	0.06362	0.00005	0.08	0.06353	0.06366	5
quadratisch- triangular	0.06359	0.00004	0.07	0.06354	0.06364	5
triangular	0.06366	7.64E-05	0.12	0.06356	0.06373	5

Die grafische Auswertung bestätigt hier wieder das Ergebnis der vorhergehenden Versuchsreihen. Die „*Mexican hat*“-artigen Funktionen schneiden deutlich schlechter ab als die „dreiecksartigen“.

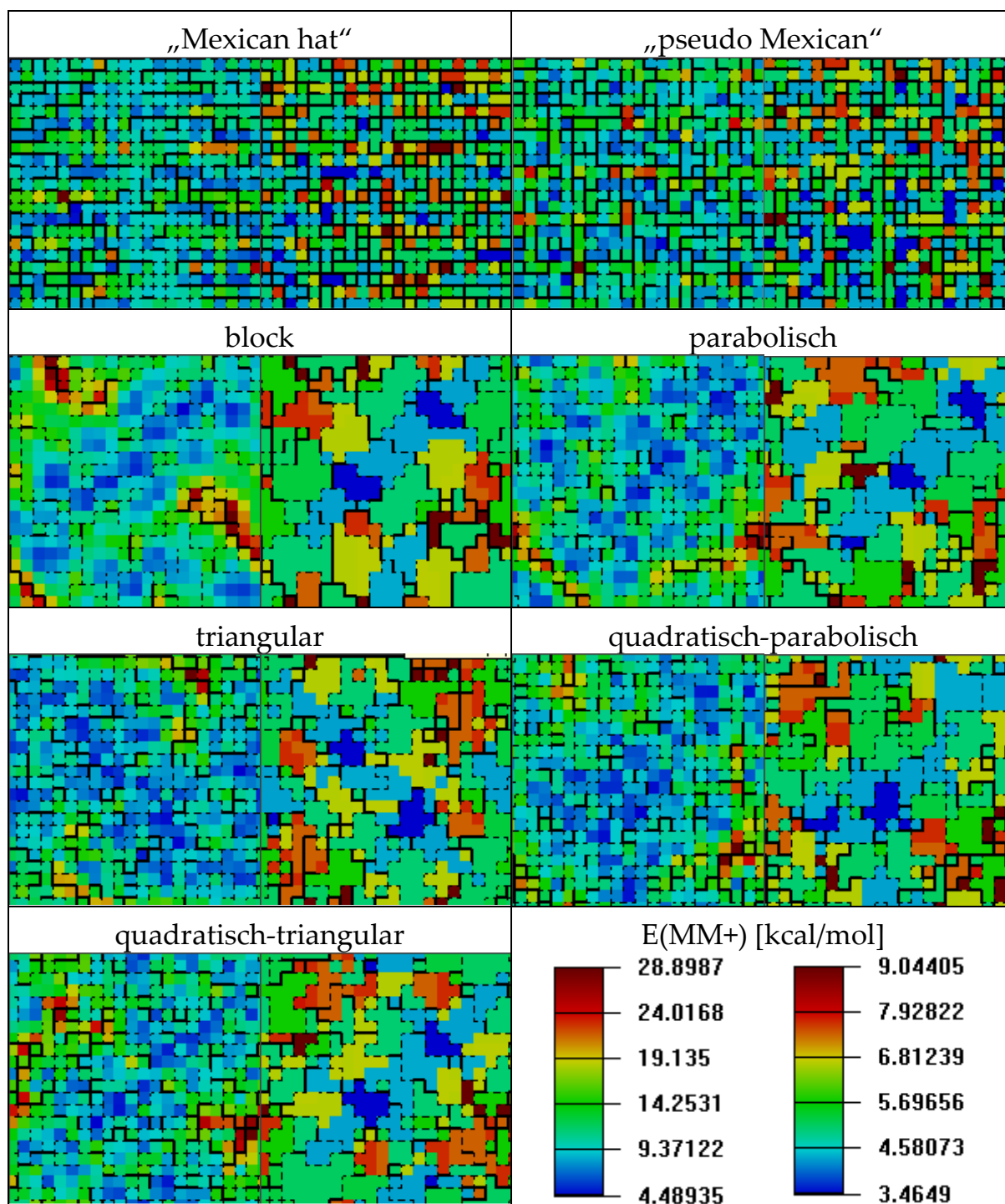




**Abbildung 87: Variation der Abstandsfunktion  $\vartheta(\zeta)$**

Die *block* Funktion zeigt zugleich den höchsten RMSDSN- und RMSDCF-Wert sowie den niedrigsten RMSG-Wert. An dieser Stelle werden MINIX1- und MINIX2-Rechnungen exemplarisch für je eine Netzkarte pro Abstandsfunktion durchgeführt, um die erhaltenen Netzkarten auf subjektive „Übersichtlichkeit“ zu überprüfen.

Die Ergebnisse sind in der Abbildung 88 dargestellt.



**Abbildung 88:** Beispielnetzkarten Gridhyperfläche Hexan – Variation der Abstandsfunktion, links partiell optimiert unter Verwendung der Netzrestraints nach COLORIX1, rechts vollständig optimiert nach COLORIX2

Es ist ersichtlich, dass die energetisch niedrigste *all-anti* Konformation in jeder Netzkarte mehrfach vorkommt. Derartige Projektionsfehler sind nicht zu vermeiden, wenn die Nachbarschaftsbeziehungen der dreidimensionalen Hyperfläche weitgehend erhalten bleiben sollen. Vom subjektiven Eindruck scheint die *block* Funktion die einfachste Netzoberfläche zu besitzen, die parabolische Funktion folgt. Andere Abstandsfunktionen zeigen relativ viele „Grenzflächenfehler“ zwischen größeren Konformationsbereichen. Der niedrigere RMSG-Wert der *block* Funktionsprojektion deckt sich ausgezeichnet

mit der in Kapitel 2.5.1 postulierten „spannungsfreien“ Projektion für niedrige RMSG-Werte.

### 2.8.3.1.7 **Fazit**

Es werden an dieser Stelle nur noch die für Hexan spezifischen Ergebnisse zusammengefasst.

- *epochs*: ermittelte minimale Trainingsschritte nach Kapitel 2.5.4  
 $MTS(RMSG, 1\%) = 6912000 = MTS(RMSDSN, 1\%)$
- $f_0$ :  
 Die nach RMSG-Wert reproduzierbarsten Kurven werden mit  $f_0 = 0.15$  erzielt. Noch niedrigere RMSG-Werte werden mit kleineren  $f_0$  Werten erzielt.
- $\Delta_{\max}(1)$   
 $\Delta_{\max}(1) = 5 \dots 10$  (50...100%) liefert reproduzierbar gute RMSG-Werte mit besseren Werten für größere  $\Delta_{\max}(1)$ . Unter  $\Delta_{\max}(1) = 5$  werden die Werte unreproduzierbar. Nach der RMSDSN-Kurve sind kleinere Werte besser als größere, jedoch kommt es bei  $\Delta_{\max}(1) < 3$  (30%) zu leicht erhöhten relativen Fehlern.
- $\eta(\xi)$   
 Die quadratische Zeitabhängigkeit zeigt deutlich bessere RMSG- und schlechtere RMSDSN-Werte. Die RMSG-Werte der quadratischen Funktion sind jedoch mit einem höheren Fehler behaftet.
- NETX / NETY  
 Ein Fehler  $> 5\%$  für RMSDSN und RMSG ergibt sich in obiger Analyse bei  $\leq 25$  Neuronen, was 3 Neuronen/Konformation entspricht.
- $\vartheta(\zeta)$   
 Die Qualitätsreihenfolge für die Funktionen der Abstandsabhängigkeit des Adaptionfaktors nach RMSG-Werten lautet (kleinere Werte sind besser):  
 „block“  $>$  „parabolisch“  $>$  „quadratisch-triangular“  $\geq$  „quadratisch-parabolisch“  $\geq$  „triangular“  $>$  „pseudo Mexican“  $>$  „Mexican hat“.  
 Die Qualitätsreihenfolge nach RMSDSN-Werten lautet (kleinere Werte sind besser):  
 „quadratisch-triangular“  $\geq$  „quadratisch-parabolisch“  $\geq$  „triangular“  $>$  „parabolisch“  $>$  „pseudo Mexican“  $>$  „Mexican hat“  $>$  „block“

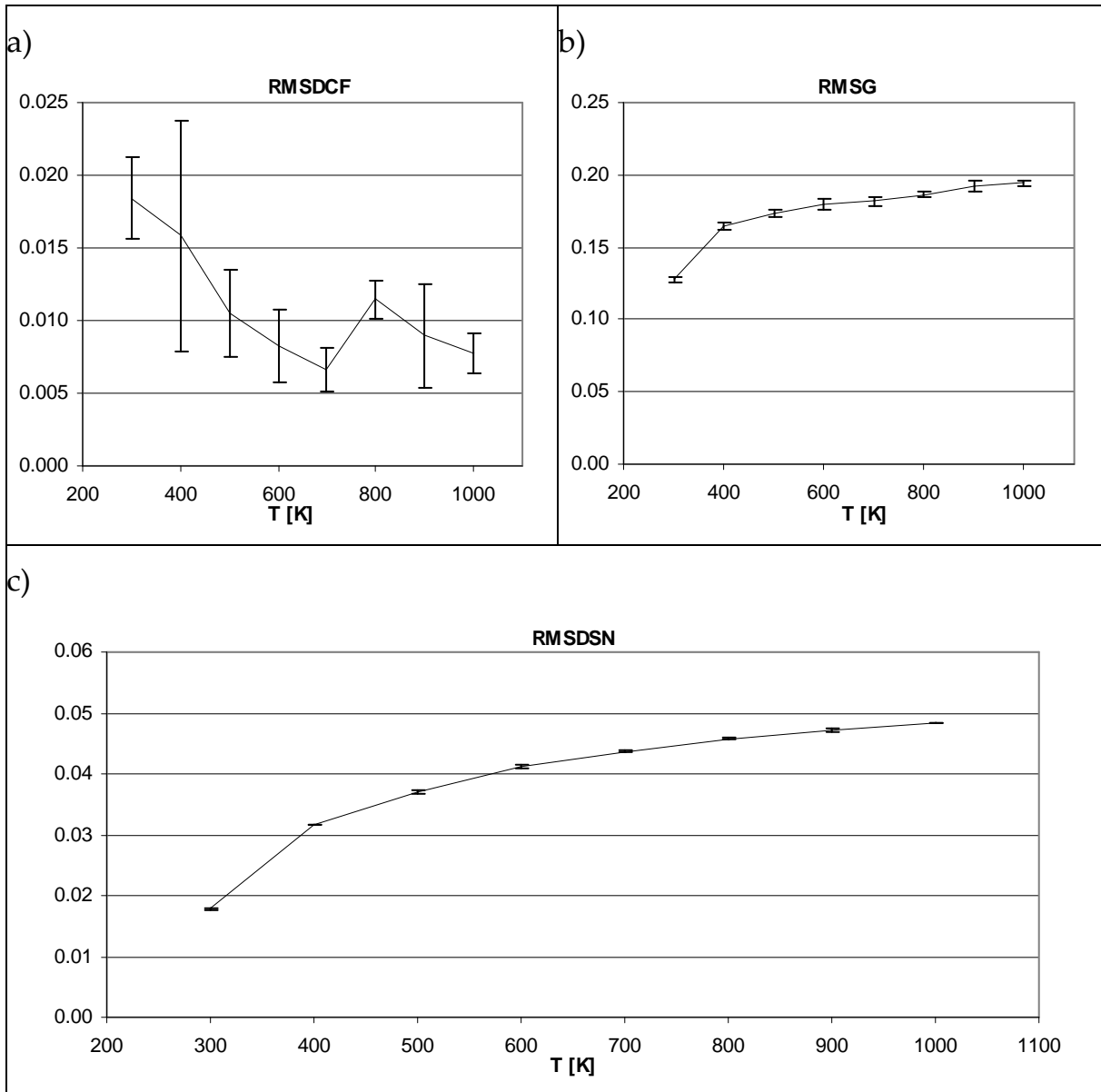
### 2.8.3.2 MD Hexan, Training der Diederwinkel

Für das Hexan-Molekül wird im Folgenden bei unterschiedlichen Temperaturen eine Gruppe Moleküldynamikrechnungen durchgeführt. (Kraftfeld HYPERCHEM: MM+, in vacuo, 40000 Strukturen, Simulationsdauer 20000 ps, Aufruf von FIX\_ROTATION.EXE bei je 2 ps, *step size* 0.001 ps, *datafreq*=1/500 *data steps*<sup>-1</sup>). Mit diesen MD-Eingabedaten werden neuronale Netze trainiert. [25x25 Neuronen; 100 *epochs*;  $\Delta_{\max}(1)=12$ ;  $f_0=0.4$ ; Abstandsfunktion parabolisch, Training der drei CC-Diederwinkel]. Die erhaltenen Qualitätsparameter sind in Tabelle 39 angeführt.

**Tabelle 39: Hexan MD-Simulationen bei unterschiedlichen Temperaturen, Training der zentralen Diederwinkel**

T [K]	Ø RMSDCF	$\sigma$	relerr%	min	max	ndata
1000	0.00770	0.00138	17.94	0.00625	0.00924	5
900	0.00896	0.00358	40.02	0.00535	0.01374	5
800	0.01148	0.00132	11.54	0.01028	0.01296	5
700	0.00662	0.00154	23.20	0.00549	0.00932	5
600	0.00829	0.00253	30.47	0.00510	0.01134	5
500	0.01054	0.00301	28.54	0.00628	0.01389	5
400	0.01581	0.00791	50.04	0.00612	0.02371	5
300	0.01844	0.00279	15.13	0.01617	0.02322	5
T [K]	Ø RMSG	$\sigma$	relerr%	min	max	ndata
1000	0.19439	0.00175	0.90	0.19258	0.19634	5
900	0.19183	0.00375	1.96	0.18761	0.19748	5
800	0.18608	0.00185	1.00	0.18433	0.18914	5
700	0.18153	0.00276	1.52	0.17935	0.18506	5
600	0.17919	0.00363	2.03	0.17497	0.18313	5
500	0.17350	0.00265	1.53	0.17088	0.17735	5
400	0.16493	0.00250	1.52	0.16244	0.16811	5
300	0.12773	0.00155	1.21	0.12575	0.12985	5
T [K]	Ø RMSDSN	$\sigma$	relerr%	min	max	ndata
1000	0.04837	0.00005	0.10	0.04829	0.04843	5
900	0.04708	0.00025	0.52	0.04683	0.04742	5
800	0.04579	0.00013	0.28	0.04561	0.04594	5
700	0.04365	0.00021	0.48	0.04336	0.04386	5
600	0.04117	0.00018	0.45	0.04101	0.04147	5
500	0.03702	0.00019	0.51	0.03683	0.03729	5
400	0.03165	0.00012	0.37	0.03154	0.03180	5
300	0.01780	0.00011	0.62	0.01771	0.01796	5

Grafisch dargestellt wird obige Tabelle in Abbildung 89. Der kontinuierlich größer werdende Konformationsraum bei höheren Temperaturen hat kontinuierlich steigende, aber reproduzierbare RMSG- und RMSDSN-Werte zur Folge.



**Abbildung 89: Variation der Temperatur – Hexan MD, Training der drei CC-Diederwinkel**

Der zunehmende Umfang des Konformationsraums bzw. die Verlagerung der „Aufenthaltswahrscheinlichkeit“ des Moleküls in höherenergetischen Konformationen lässt sich gut in Abbildung 90 nachvollziehen. Insgesamt zeigt das Molekül bereits bei niedrigeren Temperaturen eine wesentlich größere Flexibilität. Dies deutet darauf hin, dass größere Moleküle bei realistischeren Temperaturen sinnvoll simuliert werden können.



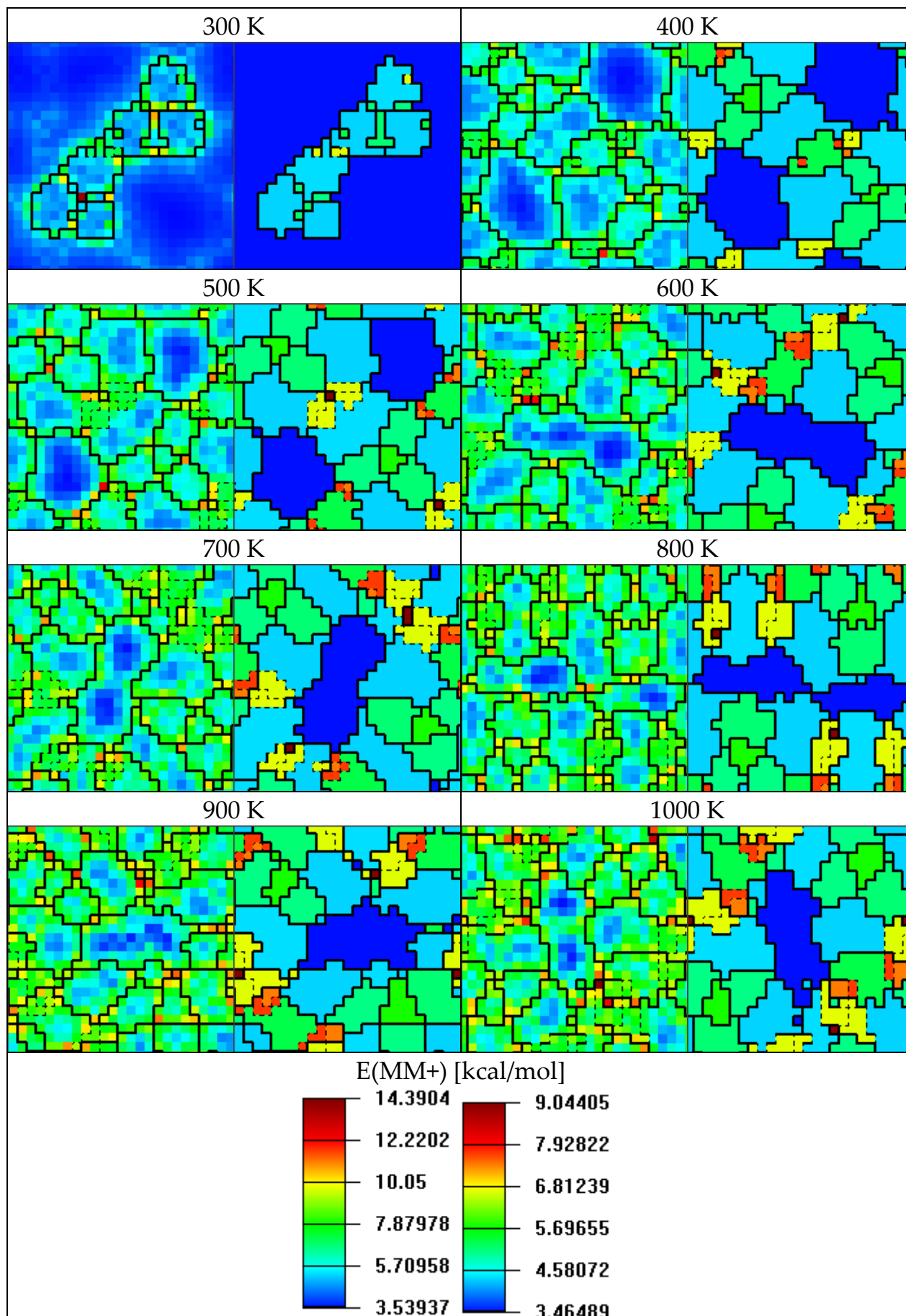


Abbildung 90: Beispielnetskarten Gridhyperfläche – Variation der Temperatur – Hexan MD, Training der CC-Diederwinkel, links partiell optimiert unter Verwendung der Netzrestraints nach COLORIX1, rechts vollständig optimiert nach COLORIX2

Die komplette Hyperfläche wird mit der 500 K - Simulation zumindest einmal abgedeckt. Eine exemplarische Auswertung dieser Netzkarte aus Abbildung 90 ist in Abbildung 81 durchgeführt.

### 2.8.3.3 MD Hexan, Training 30 zufälliger Distanzen

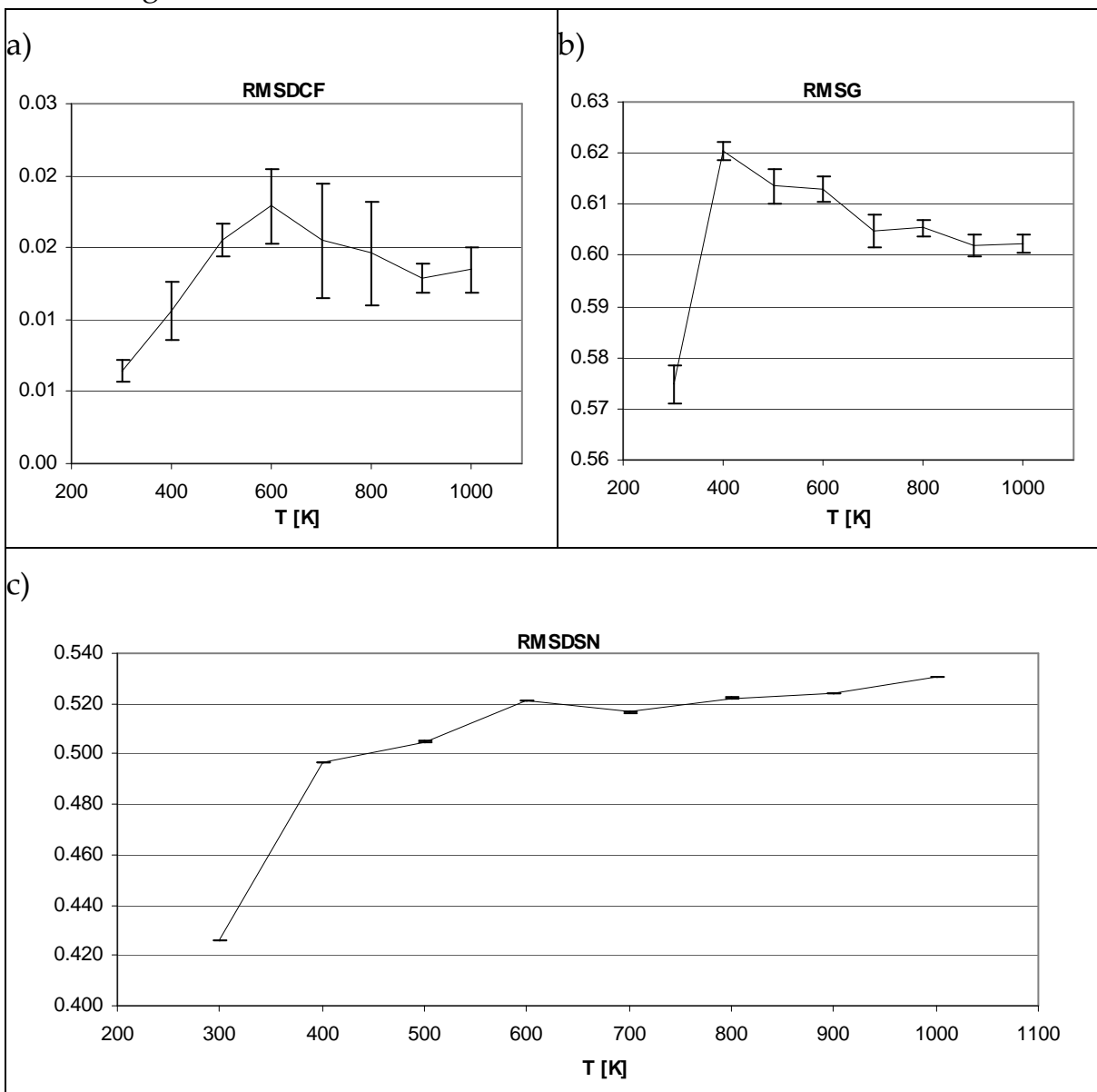
Das Training aller C-C, C-H und H-H Abstände ist beim Hexan rechenzeitmäßig bereits unpraktikabel, statt dessen werden dreißig zufällig ausgewählte C-C, C-H und H-H Abstände (die doppelte Anzahl aller Schweratomsabstände) trainiert. Wie bei den vorherigen Molekülen wird dabei die Hyperfläche zusätzlich kompliziert durch Numerationsisomere. Als Eingabedaten werden die gleichen MD-Rechnungen wie im vorherigen Kapitel verwendet.

Die statistischen Daten des Netztrainings sind in Tabelle 40 aufgeführt. [25x25 Neuronen; 20 epochs;  $\Delta_{\max}(1)=12$ ;  $f_0=0.4$ ; Abstandsfunktion parabolisch, Training 30 zufällig ausgewählter Distanzen].

**Tabelle 40:** Hexan MD-Simulationen bei unterschiedlichen Temperaturen, Training 30 zufällig ausgewählter Distanzen

T [K]	Ø RMSDCF	$\sigma$	relerr%	min	max	ndata
1000	0.01346	0.00159	11.83	0.01089	0.01525	5
900	0.01284	0.00099	7.72	0.01145	0.01421	5
800	0.01463	0.00360	24.62	0.01097	0.01925	5
700	0.01550	0.00398	25.67	0.01033	0.01887	5
600	0.01787	0.00259	14.47	0.01456	0.02126	5
500	0.01557	0.00116	7.44	0.01417	0.01731	5
400	0.01061	0.00198	18.63	0.00768	0.01308	5
300	0.00643	0.00072	11.23	0.00579	0.00764	5
T [K]	Ø RMSG	$\sigma$	relerr%	min	max	ndata
1000	0.60218	0.00182	0.30	0.59969	0.60442	5
900	0.60202	0.00210	0.35	0.59888	0.60403	5
800	0.60531	0.00166	0.27	0.60376	0.60784	5
700	0.60477	0.00327	0.54	0.60065	0.60910	5
600	0.61303	0.00245	0.40	0.60995	0.61672	5
500	0.61360	0.00339	0.55	0.60903	0.61817	5
400	0.62041	0.00168	0.27	0.61839	0.62298	5
300	0.57475	0.00360	0.63	0.56893	0.57785	5
T [K]	Ø RMSDSN	$\sigma$	relerr%	min	max	ndata
1000	0.53050	0.00015	0.03	0.53031	0.53063	5
900	0.52414	0.00011	0.02	0.52400	0.52430	5
800	0.52225	0.00027	0.05	0.52190	0.52253	5
700	0.51669	0.00029	0.06	0.51624	0.51700	5
600	0.52128	0.00019	0.04	0.52108	0.52146	5
500	0.50491	0.00011	0.02	0.50478	0.50505	5
400	0.49684	0.00014	0.03	0.49663	0.49701	5
300	0.42597	0.00030	0.07	0.42565	0.42642	5

Der RMSDSN-Wert steigt kontinuierlich an, was mit der kontinuierlichen Vergrößerung des abzubildenden Konformationsraums korreliert. Der RMSDCF-Wert erreicht bei 600 K ein Maximum, der RMSG-Wert bei 400 K. Beide sinken bei noch höheren Temperaturen dann kontinuierlich ab, wie in Abbildung 91 zu sehen ist.



**Abbildung 91: Variation der Temperatur – Hexan MD, Training 30 zufällig ausgewählter Distanzen**

Für eine Erklärung der RMSG-Werte können die Beispielnetskarten aus Abbildung 92 herangezogen werden. Bei 400 K kommt es zu einem Wechsel der Netztopologie. Mit steigender Temperatur verteilen sich die Konformationen gleichmäßiger über die Netzoberfläche.

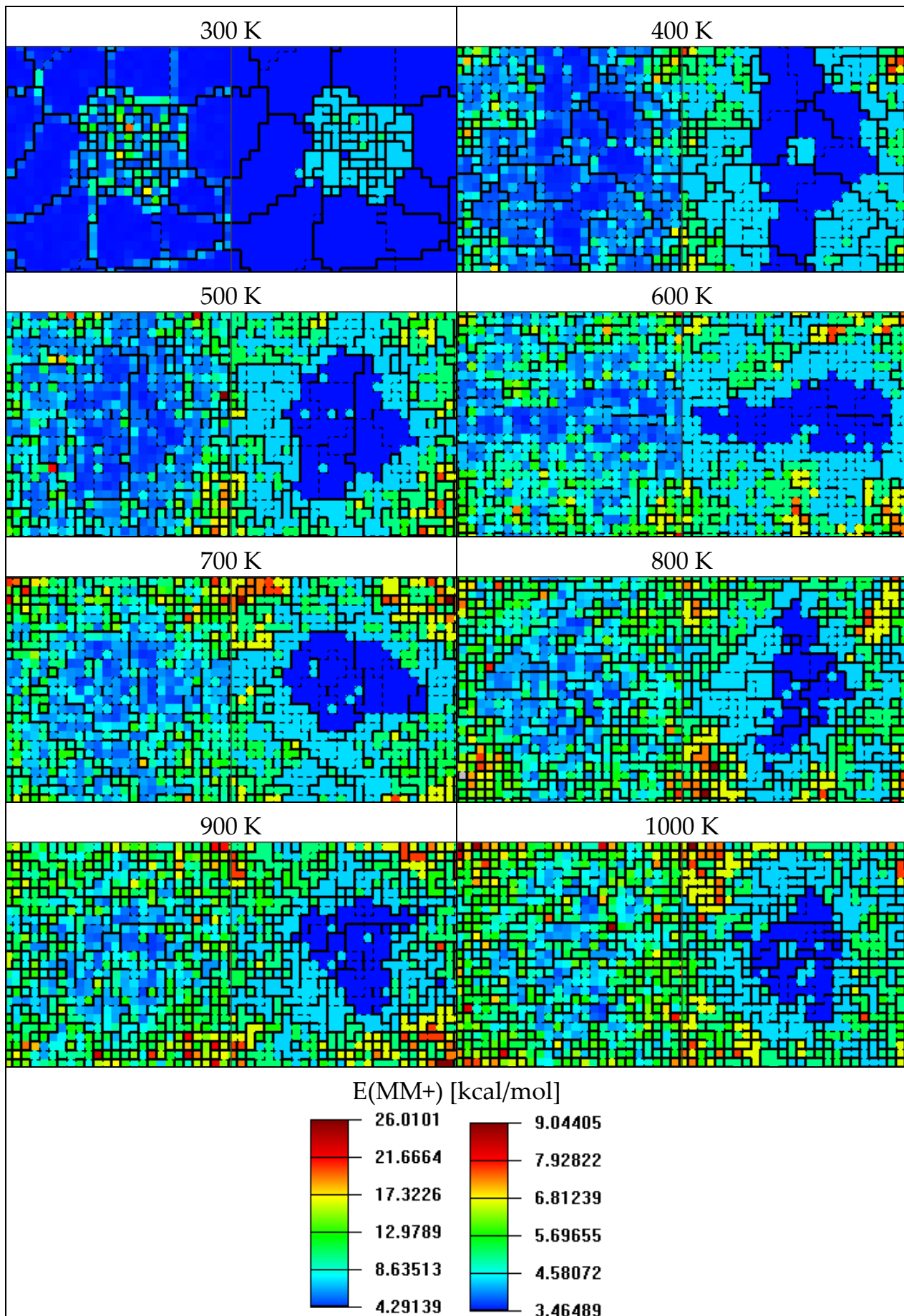


Abbildung 92: Beispielnetzarten Gridhyperfläche – Variation der Temperatur – Hexan MD, Training 30 zufällig ausgewählter Distanzen, links partiell optimiert unter Verwendung der Netzrestraints nach COLORIX1, rechts vollständig optimiert nach COLORIX2

Interessanterweise sind die Netzoberflächen trotz der stärkeren Strukturierungen auf Grund der Numerationsisomerie einfach auszuwerten mit klar abgetrennten Bereichen zwischen Niedrig- und Hochenergiekonformationen.

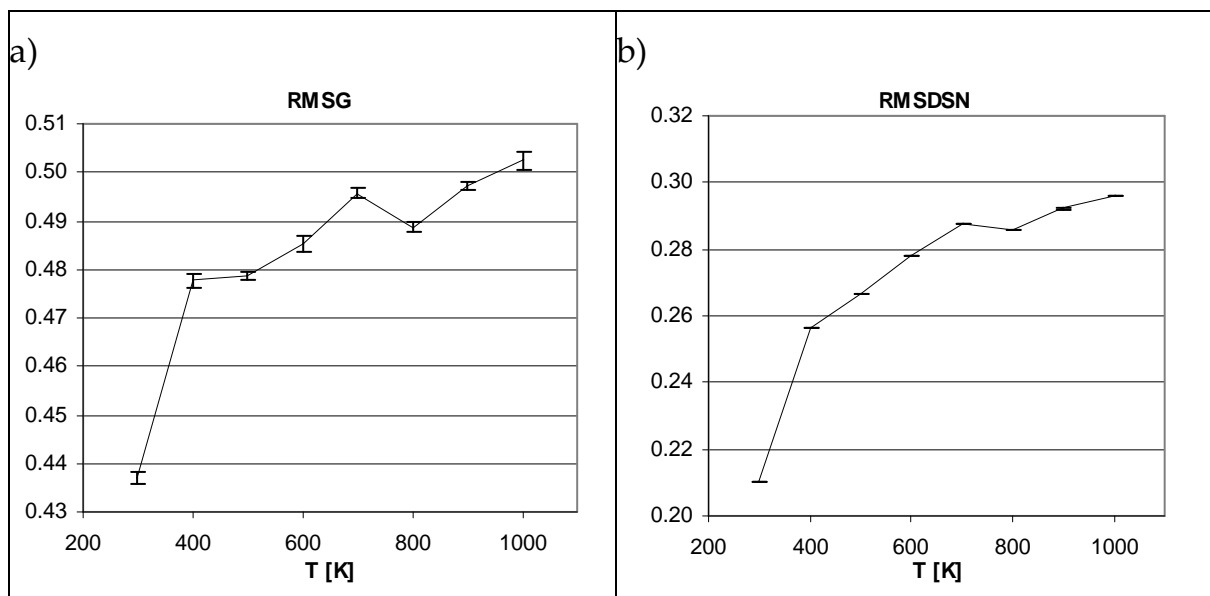
### 2.8.3.4 MD Hexan, Training der Schweratomdistanzen

Das Training der Schweratomsmatrix führt wiederum zu einer vereinfachten Hyperfläche. Die statistischen Daten des Netztrainings [25x25 Neuronen; 100 epochs;  $\Delta_{\max}(1)=12$ ;  $f_0=0.4$ ; Abstandsfunktion parabolisch, Training der CC-Distanzen] sind in Tabelle 41 aufgeführt.

**Tabelle 41: Hexan MD-Simulationen bei unterschiedlichen Temperaturen, Training der Schweratomdistanzmatrix**

T [K]	Ø RMSG	$\sigma$	relerr%	min	max	ndata
1000	0.50237	0.00188	0.37	0.50072	0.50523	5
900	0.49704	0.00083	0.17	0.49569	0.49779	5
800	0.48872	0.00103	0.21	0.48739	0.48983	5
700	0.49560	0.00102	0.21	0.49427	0.49681	5
600	0.48522	0.00155	0.32	0.48346	0.48713	5
500	0.47859	0.00088	0.18	0.47745	0.47950	5
400	0.47771	0.00135	0.28	0.47602	0.47933	5
300	0.43714	0.00125	0.29	0.43589	0.43882	5
T [K]	Ø RMSDSN	$\sigma$	relerr%	min	max	ndata
1000	0.29602	0.00009	0.03	0.29589	0.29611	5
900	0.29225	0.00016	0.05	0.29207	0.29243	5
800	0.28589	0.00009	0.03	0.28581	0.28604	5
700	0.28739	0.00004	0.01	0.28734	0.28745	5
600	0.27803	0.00012	0.04	0.27791	0.27816	5
500	0.26662	0.00009	0.03	0.26649	0.26671	5
400	0.25655	0.00008	0.03	0.25646	0.25663	5
300	0.21018	0.00008	0.04	0.21011	0.21031	5

Die Graphen der Trainings-Qualitätsparameter der CC-Distanzen folgen vom Trend eher den Graphen der Diederwinkel als den Graphen der Matrix der zufällig ausgewählten Abstände. So steigen die RMSG- und RMSDSN-Werte mit der Temperatur der MD-Werte.



**Abbildung 93:** Variation der Temperatur – Hexan MD, Training der Schweratomdistanzmatrix

Die Topologie von beispielhaften Netzoberflächen bei verschiedenen Temperaturen zeigt näherungsweise eine dreifache Symmetrieachse, wobei die dunkelblauen Konformationsbereiche der *all-anti* Konformation (aaa) entspricht und die drei türkisfarbenen Konformationsbereiche die Konformationen mit exakt einem *gauche* CC-Diederwinkel darstellen, wobei zwischen *gauche*<sup>+</sup> (G) und *gauche*<sup>-</sup> (g) nicht unterschieden werden kann.

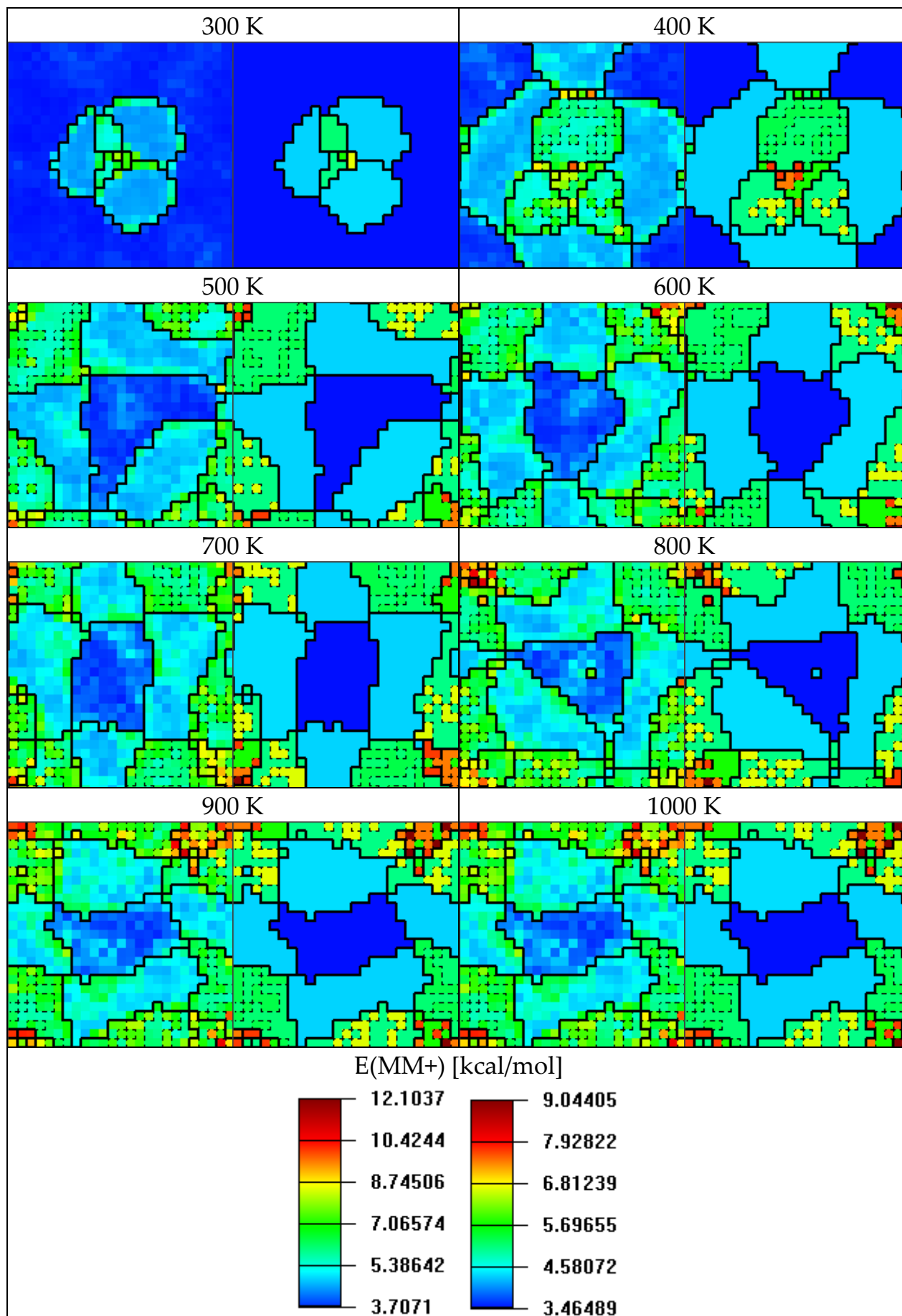


Abbildung 94: Beispielnetskarten Gridhyperfläche – Variation der Temperatur – Hexan MD, Training der Schweratomdistanzmatrix, links partiell optimiert unter Verwendung der Netzrestraints nach COLORIX1, rechts vollständig optimiert nach COLORIX2

## 2.8.4 n-Heptan

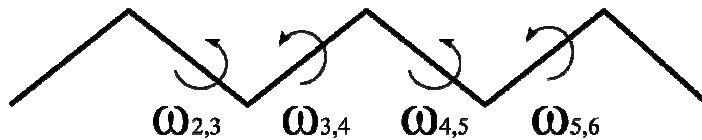


Abbildung 95: CC-Diederwinkel im n-Heptan

Das n-Heptanmolekül enthält vier frei rotierbare Diederwinkel, wenn man die freie Drehbarkeit der terminalen Methylgruppen außer Acht lässt. Die Topologie der Hyperfläche ist damit mindestens 4-dimensional und kann an dieser Stelle nicht mehr sinnvoll als klassisches *grid*-Diagramm präsentiert werden, wie es noch beim Butan, Pentan und eingeschränkt beim Hexan möglich war. Die Anzahl der Konformationen kann mit  $3^4=81$  abgeschätzt werden.

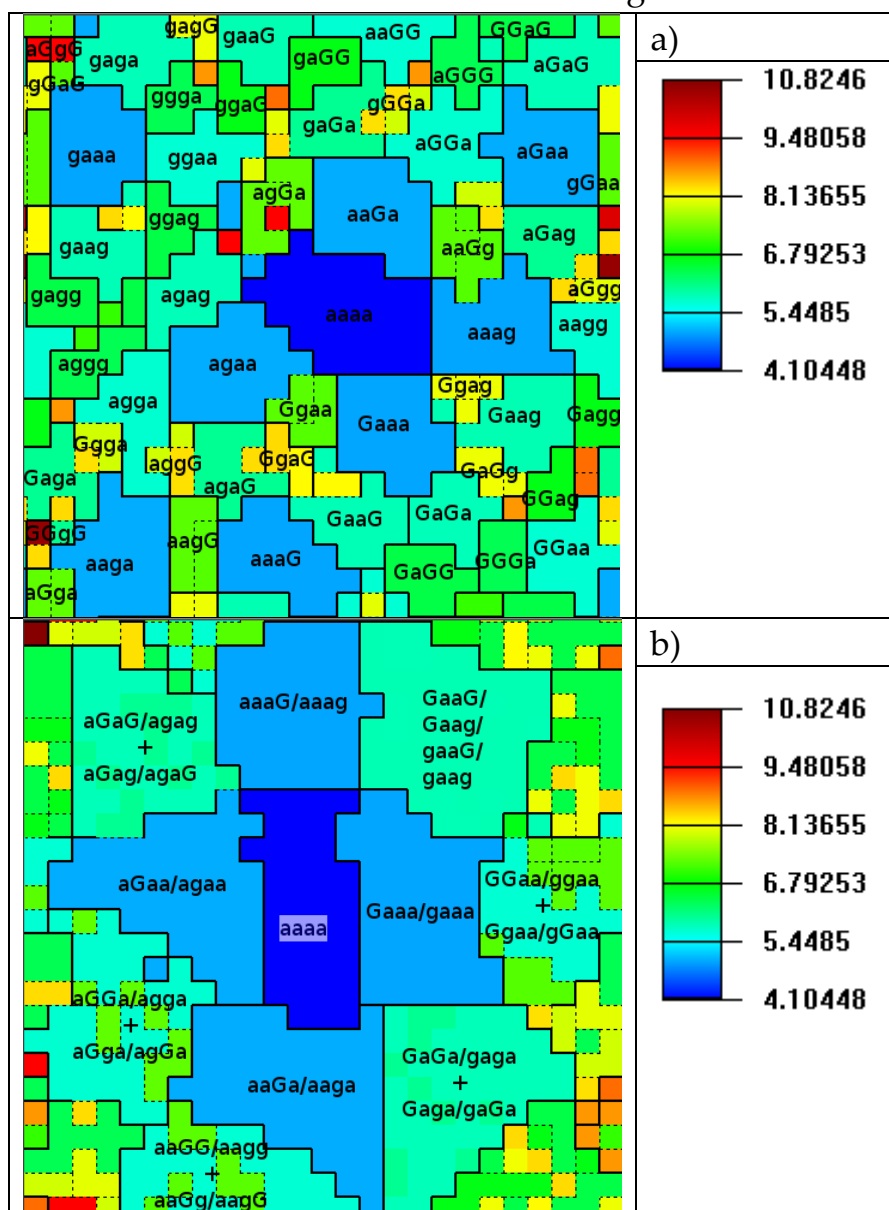


Abbildung 96: Beispielprojektion der MM+ Hyperfläche des Heptans und Energieskala in kcal/mol

a) Training der Diederwinkel, MD 600K

b) Training der Schweratommatrix, MD 700K



In Abbildung 96 werden exemplarische Netzkarten aus Moleküldynamikdaten mit Hilfe des Skriptes PRINT\_SELECTION\_VALUES\_GUI.TK ausgewertet. Je nach verwendetem Trainingsatz können dabei Enantiomere unterschieden werden, siehe a), oder auch nicht, siehe b).

#### 2.8.4.1 Training der Grid-Hyperfläche

Eine *grid*-Rechnung über die in Abbildung 95 angegebenen Diederwinkel kann für den folgenden Teil immer noch durchgeführt werden. Variiert werden mit dem Programm GRID.TCL (5.2.2) unter Verwendung des Kraftfelds MM+ (HYPERCHEM) folgende CC-Diederwinkel in 30°-Schritten:

- $\omega_{2,3}$ : 0, 30, 60, ..., 330
- $\omega_{3,4}$ : 0, 30, 60, ..., 330
- $\omega_{4,5}$ : 0, 30, 60, ..., 330
- $\omega_{5,6}$ : 0, 30, 60, ..., 330

Die generierten Strukturen (insgesamt 20736) werden wie beim Butan, Pentan und Hexan geometrieoptimiert, wobei besagte Diederwinkel während der Optimierung eingefroren bleiben.

Für die Ermittlung der RMSDCF-Werte wird eine Klassifizierung der Heptan-konformationen mittels des Programms CLASSIFY\_SNP\_CONFORMATIONS.EXE durchgeführt, siehe Kapitel 5.3.5.

Trainiert werden die Kohonennetze in den folgenden Unterkapiteln mit Vorgabewerten von  $f_0=0.15$ ,  $epochs=1000$ ,  $neighborhood\ function =$  parabolisch,  $NETX=NETY=21$ ,  $linear\ learning\ decay$  für die nichtvariieren Netzparameter.

Das Heptan-Molekül zeigt eine synthetische Hyperfläche, welche komplex genug aber nicht zu komplex ist, um der Frage nachzugehen, welche Qualitätsparameter im Zweifelsfalle bevorzugt als Benchmark herhalten sollen, um möglichst übersichtliche Karten zu erhalten.

Für die Parameter  $f_0$ ,  $\Delta_{max}(1)$ ,  $\eta(\xi)$ , und die Abstandsfunktion werden beispielhafte Netzkarten dem MINIX1- und MINIX2-Protokoll unterworfen und subjektiv beurteilt.

### 2.8.4.1.1 Variation der epochs

Tabelle 42: Heptan - Variation epochs

epochs	Ø RMSDCF	$\sigma$	relerr%	min	max	ndata
1000	0.00348	0.00018	5.26	0.00327	0.00365	5
500	0.00322	0.00017	5.24	0.00302	0.00342	5
300	0.00334	0.00021	6.30	0.00299	0.00354	5
200	0.00318	0.00017	5.36	0.00300	0.00342	5
100	0.00383	0.00018	4.62	0.00355	0.00402	5
50	0.00396	0.00033	8.25	0.00364	0.00450	5
10	0.00548	0.00028	5.19	0.00512	0.00588	5
2	0.00738	0.00078	10.52	0.00639	0.00823	5
1	0.00872	0.00049	5.57	0.00807	0.00934	5
epochs	Ø RMSG	$\sigma$	relerr%	min	max	ndata
1000	0.38567	0.00116	0.30	0.38435	0.38709	5
500	0.38495	0.00194	0.50	0.38335	0.38818	5
300	0.38102	0.00186	0.49	0.37993	0.38432	5
200	0.37796	0.00124	0.33	0.37626	0.37959	5
100	0.35864	0.00053	0.15	0.35813	0.35928	5
50	0.32871	0.00101	0.31	0.32696	0.32943	5
10	0.27715	0.00085	0.31	0.27581	0.27813	5
2	0.25815	0.00845	3.27	0.25099	0.27022	5
1	0.24781	0.00954	3.85	0.23472	0.25788	5
epochs	Ø RMSDSN	$\sigma$	relerr%	min	max	ndata
1000	0.12355	0.00008	0.06	0.12346	0.12364	5
500	0.12389	0.00005	0.04	0.12383	0.12395	5
300	0.12444	0.00014	0.11	0.12425	0.12459	5
200	0.12511	0.00011	0.09	0.12496	0.12527	5
100	0.12763	0.00006	0.05	0.12758	0.12773	5
50	0.13349	0.00037	0.28	0.13317	0.13409	5
10	0.15064	0.00038	0.25	0.15014	0.15099	5
2	0.16711	0.00128	0.76	0.16518	0.16843	5
1	0.17673	0.00123	0.70	0.17498	0.17818	5

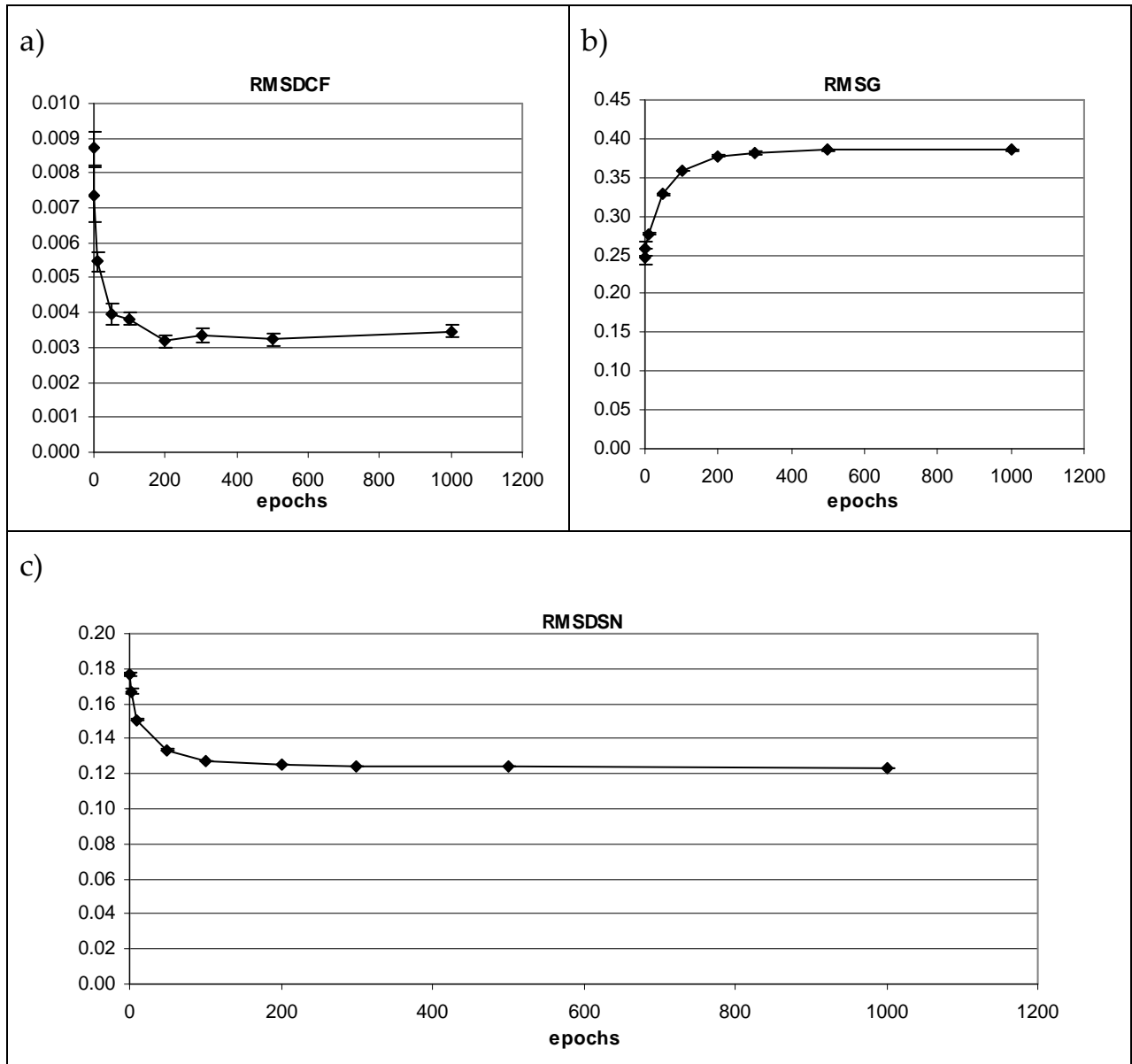
Eine grafische Auftragung der Qualitätsparameter wird in Abbildung 97 durchgeführt. Der Trend der Kurven ist praktisch identisch zum Hexan, so dass sich eine weitere Diskussion der Ergebnisse erübrigt. Per Extrapolation für  $epochs > 10$  werden  $RMSG_{\text{end}}^{[107]}$  und  $RMSDSN_{\text{end}}^{[108]}$  mit dem Programm Origin<sup>[105]</sup> ermittelt.

$$RMSG_{\text{end}} = 0.38683 \pm 0.00082, \quad \text{epoch}_{\text{end}}(\text{RMSG}) = 500$$

$$RMSDSN_{\text{end}} = 0.12351 \pm 0.00009, \quad \text{epoch}_{\text{end}}(\text{RMSDSN}) = 300$$

$$\text{MTS}(\text{RMSG}, 1\%) = 20736 * 500 = 10368000$$

$$\text{MTS}(\text{RMSDSN}, 1\%) = 20736 * 300 = 6220800$$

Abbildung 97: Variation der *epochs*, a) RMSDCF, b) RMSG, c) RMSDSN

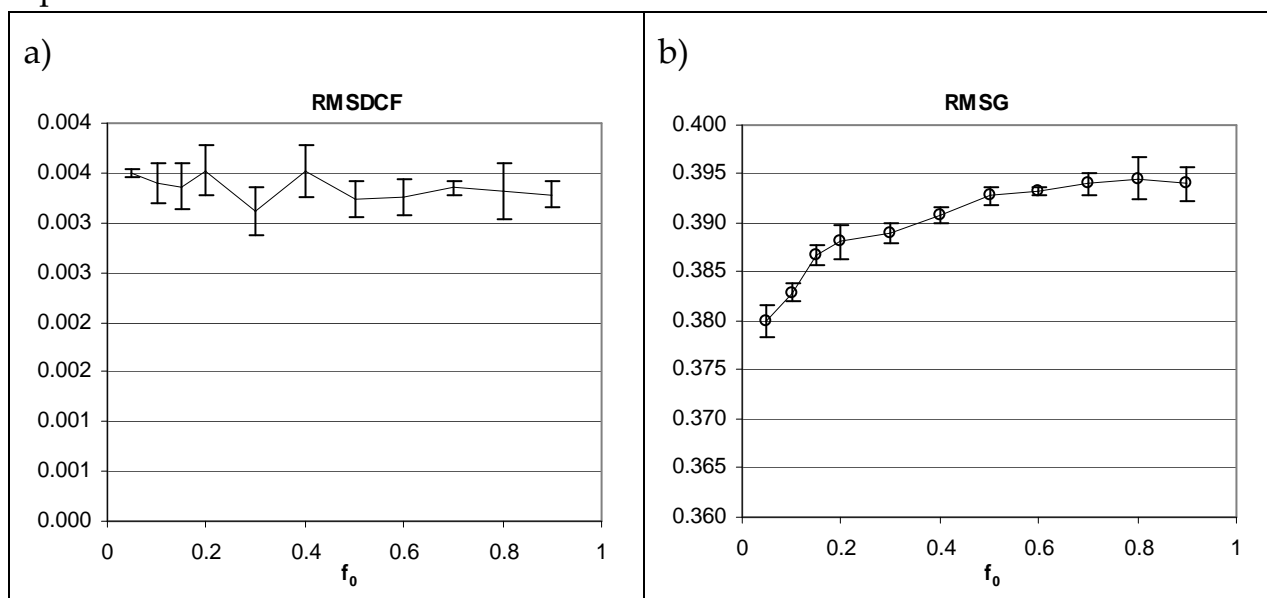
### 2.8.4.1.2 Variation von $f_0$

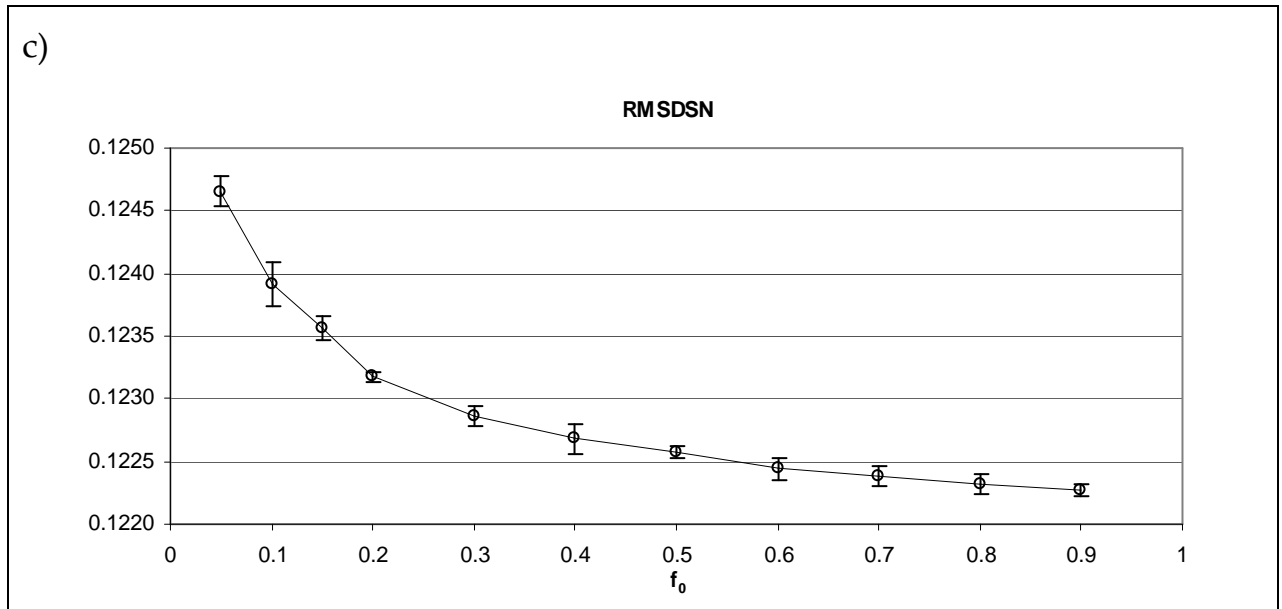
Tabelle 43: Variation des maximalen Adaptionfaktors  $f_0$  nach Gleichung (2.12)

$f_0$	$\bar{\sigma}$ RMSDCF	$\sigma$	relerr%	min	max	ndata
0.05	0.00350	0.00005	1.30	0.00343	0.00355	5
0.1	0.00340	0.00020	5.97	0.00317	0.00370	5
0.15	0.00336	0.00023	6.81	0.00311	0.00373	5
0.2	0.00352	0.00025	7.10	0.00322	0.00371	5
0.25	0.00312	0.00024	7.82	0.00287	0.00351	5
0.3	0.00351	0.00026	7.43	0.00316	0.00386	5
0.4	0.00324	0.00018	5.48	0.00302	0.00343	5
0.5	0.00326	0.00018	5.49	0.00308	0.00352	5
0.6	0.00335	0.00007	2.17	0.00325	0.00345	5
0.7	0.00332	0.00029	8.63	0.00299	0.00368	5
0.8	0.00328	0.00012	3.80	0.00313	0.00347	5
0.9	0.00350	0.00005	1.30	0.00343	0.00355	5

$f_0$	$\bar{\sigma}$ RMSG	$\sigma$	relerr%	min	max	ndata
0.05	0.37992	0.00161	0.42	0.37797	0.38235	5
0.1	0.38292	0.00094	0.25	0.38193	0.38397	5
0.15	0.38675	0.00099	0.26	0.38563	0.38798	5
0.2	0.38807	0.00180	0.46	0.38563	0.38989	5
0.25	0.38899	0.00109	0.28	0.38791	0.39054	5
0.3	0.39080	0.00082	0.21	0.38956	0.39171	5
0.4	0.39280	0.00094	0.24	0.39196	0.39429	5
0.5	0.39321	0.00041	0.10	0.39264	0.39363	5
0.6	0.39401	0.00110	0.28	0.39252	0.39515	5
0.7	0.39457	0.00221	0.56	0.39269	0.39725	5
0.8	0.39400	0.00167	0.42	0.39138	0.39537	5
0.9	0.37992	0.00161	0.42	0.37797	0.38235	5
$f_0$	$\bar{\sigma}$ RMSDSN	$\sigma$	relerr%	min	max	ndata
0.05	0.12465	0.00012	0.10	0.12451	0.12483	5
0.1	0.12391	0.00018	0.14	0.12380	0.12422	5
0.15	0.12356	0.00009	0.08	0.12346	0.12372	5
0.2	0.12317	0.00004	0.04	0.12312	0.12324	5
0.25	0.12286	0.00007	0.06	0.12275	0.12295	5
0.3	0.12268	0.00012	0.10	0.12258	0.12285	5
0.4	0.12258	0.00005	0.04	0.12250	0.12264	5
0.5	0.12244	0.00009	0.07	0.12234	0.12257	5
0.6	0.12238	0.00009	0.07	0.12230	0.12252	5
0.7	0.12232	0.00008	0.07	0.12221	0.12243	5
0.8	0.12227	0.00005	0.04	0.12218	0.12231	5
0.9	0.12465	0.00012	0.10	0.12451	0.12483	5

Die graphisch aufgetragenen Werte in Abbildung 98 zeigen wiederum einen praktisch identischen Trend wie beim Hexan.





**Abbildung 98:** Variation von  $f_0$ , a) RMSDCF, b) RMSG, c) RMSDSN

Wie bereits beim Hexan, soll am Beispiel des Heptans der visuelle Einfluss der unterschiedlichen Qualitätsparameter auf die Netzarten untersucht werden. Zu diesem Zweck werden in Abbildung 99 einzelne Netzfunktion exemplarisch den MINIX1/COLORIX1 und MINIX2/COLORIX2 Modulen zugeführt. Die erhaltenen Netzarten unterscheiden sich nicht deutlich genug, um einen eindeutigen Trend auszumachen. Dies stimmt überein mit den geringen absoluten Varianzen in den RMSG- und RMSDSN-Werten.

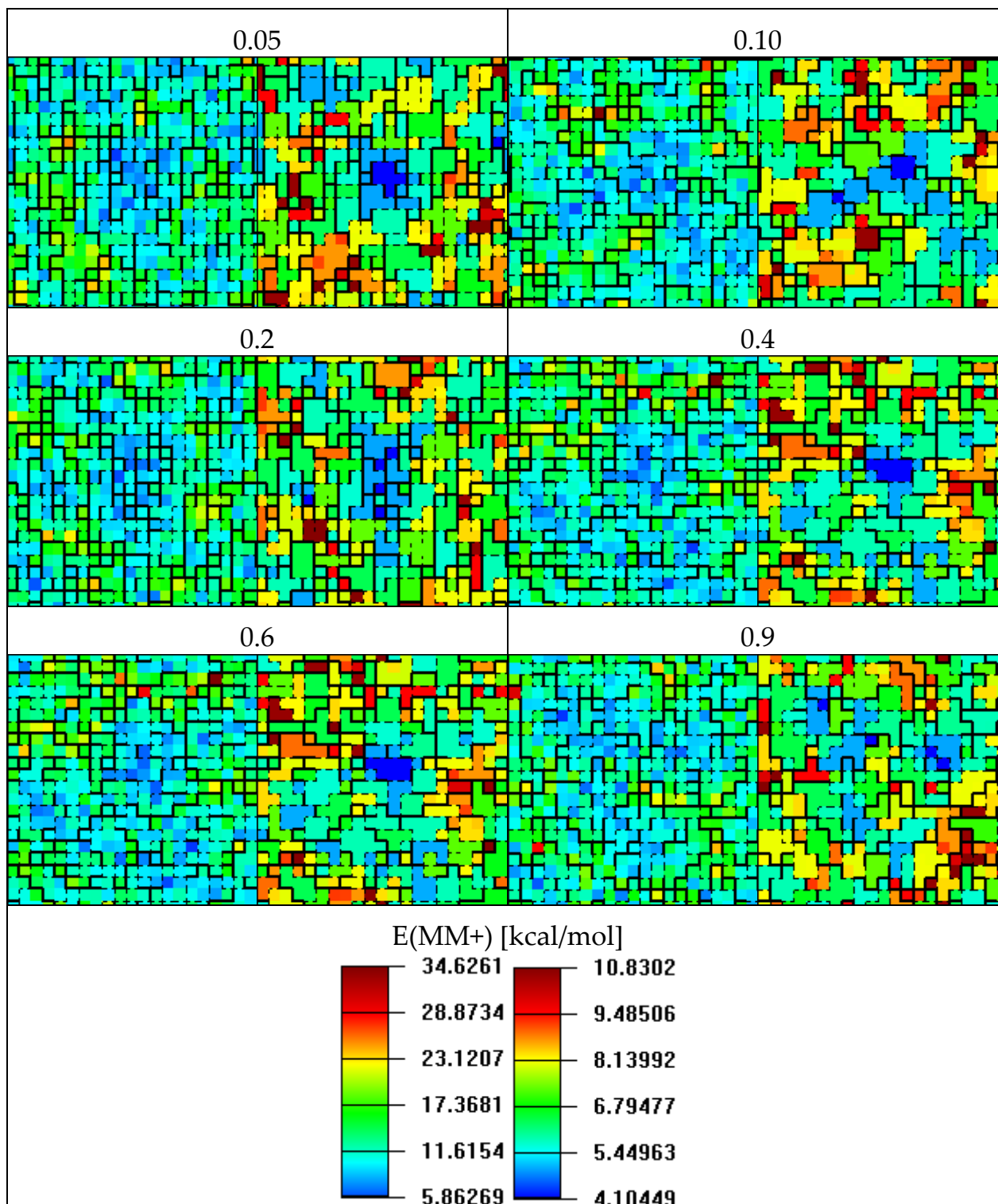
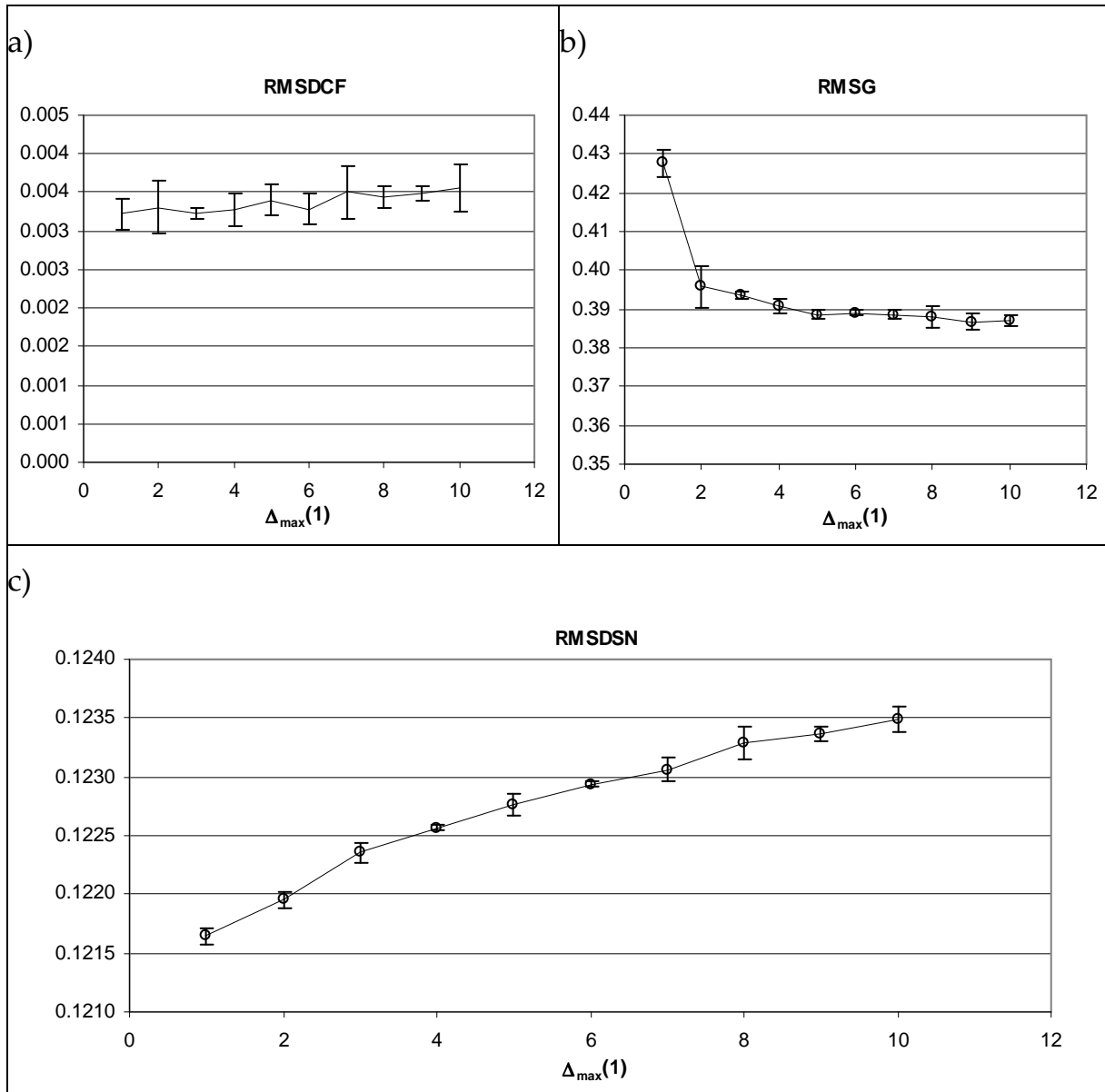


Abbildung 99: Beispielnetzarten Gridhyperfläche – Variation von  $f_0$ , Heptan, links partiell optimiert unter Verwendung der Netzrestraints nach COLORIX1, rechts vollständig optimiert nach COLORIX2

2.8.4.1.3 Variation von  $\Delta_{\max}(1)$ Tabelle 44: Variation von  $\Delta_{\max}(1)$  nach Gleichung (2.15)

$\Delta_{\max}(1)$	$\bar{\sigma}$ RMSDCF	$\sigma$	relerr%	min	max	ndata
10	0.00356	0.00031	8.75	0.00324	0.00390	5
9	0.00349	0.00010	2.95	0.00337	0.00363	5
8	0.00344	0.00015	4.28	0.00322	0.00356	5
7	0.00350	0.00035	10.05	0.00313	0.00396	5
6	0.00328	0.00020	6.15	0.00300	0.00351	5
5	0.00340	0.00020	5.83	0.00310	0.00355	5
4	0.00327	0.00021	6.42	0.00308	0.00362	5
3	0.00323	0.00007	2.31	0.00313	0.00333	5
2	0.00330	0.00034	10.42	0.00276	0.00361	5
1	0.00322	0.00020	6.11	0.00305	0.00355	5
$\Delta_{\max}(1)$	$\bar{\sigma}$ RMSG	$\sigma$	relerr%	min	max	ndata
10	0.38703	0.00139	0.36	0.38536	0.38851	5
9	0.38678	0.00218	0.56	0.38305	0.38881	5
8	0.38792	0.00288	0.74	0.38455	0.39226	5
7	0.38860	0.00130	0.33	0.38706	0.39018	5
6	0.38912	0.00084	0.22	0.38849	0.39056	5
5	0.38862	0.00130	0.33	0.38640	0.38976	5
4	0.39098	0.00184	0.47	0.38839	0.39317	5
3	0.39364	0.00086	0.22	0.39245	0.39459	5
2	0.39585	0.00536	1.35	0.38743	0.40175	5
1	0.42763	0.00346	0.81	0.42317	0.43257	5
$\Delta_{\max}(1)$	$\bar{\sigma}$ RMSDSN	$\sigma$	relerr%	min	max	ndata
10	0.12349	0.00010	0.08	0.12332	0.12360	5
9	0.12337	0.00006	0.05	0.12332	0.12346	5
8	0.12329	0.00014	0.11	0.12316	0.12350	5
7	0.12306	0.00010	0.08	0.12296	0.12321	5
6	0.12294	0.00003	0.02	0.12291	0.12297	5
5	0.12276	0.00009	0.08	0.12265	0.12289	5
4	0.12257	0.00003	0.02	0.12252	0.12260	5
3	0.12236	0.00008	0.07	0.12227	0.12246	5
2	0.12195	0.00007	0.06	0.12188	0.12204	5
1	0.12165	0.00007	0.06	0.12156	0.12172	5

Die grafische Auftragung obiger Tabellenwerte ergibt wiederum einen ähnlichen Trend wie beim Pentantraining. Als interessante Abweichung gibt es einen „Knick“ in der RMSG-Kurve bei  $\Delta_{\max}(1)=5$  (50%). Dies ermöglicht die Angabe dieses Wertes als Kompromisswert bei der gleichzeitigen Optimierung der RMSG- und RMSDSN-Werte. Ein ähnlicher Effekt war bereits beim Butan zu sehen.



**Abbildung 100: Variation von  $\Delta_{\max}(1)$ , a) RMSDCF, b) RMSG, c) RMSDSN**

Auch der  $\Delta_{\max}(1)$ -Parameter ist einer der Parameter, welcher antagonistisches Verhalten der RMSG- und RMSDSN-Qualitätsparameter zeigt. Beispielnetskarten nach MINIX1/COLORIX1 und MINIX2/COLORIX2 sind in Abbildung 101 wiedergegeben. Auch hier sind die absoluten Varianzen von RMSG und RMSDSN für  $\Delta_{\max}(1) > 2$  sehr gering und es werden keine offensichtlichen Unterschiede in der Qualität der erhaltenen Netskarten offenbar.



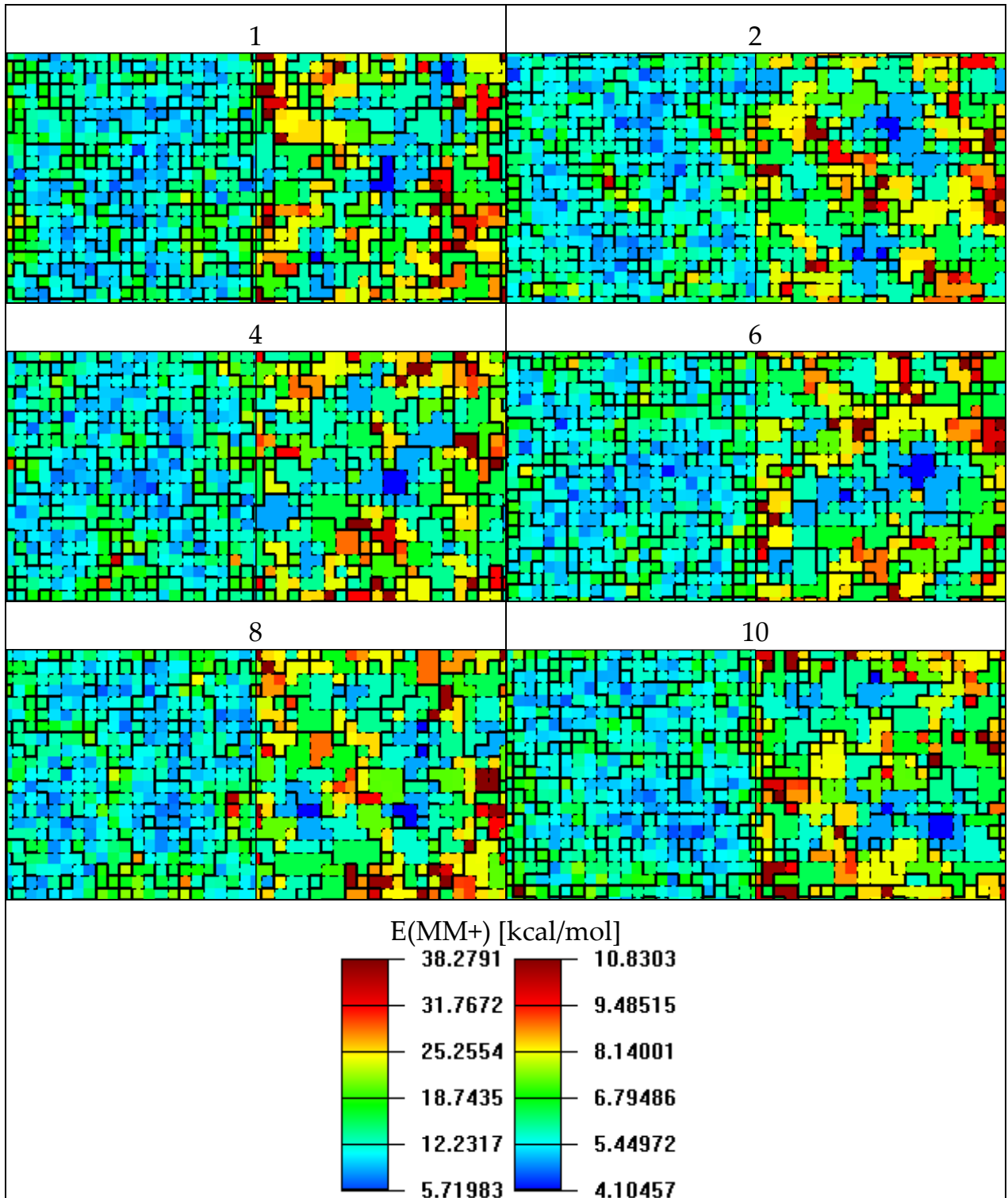


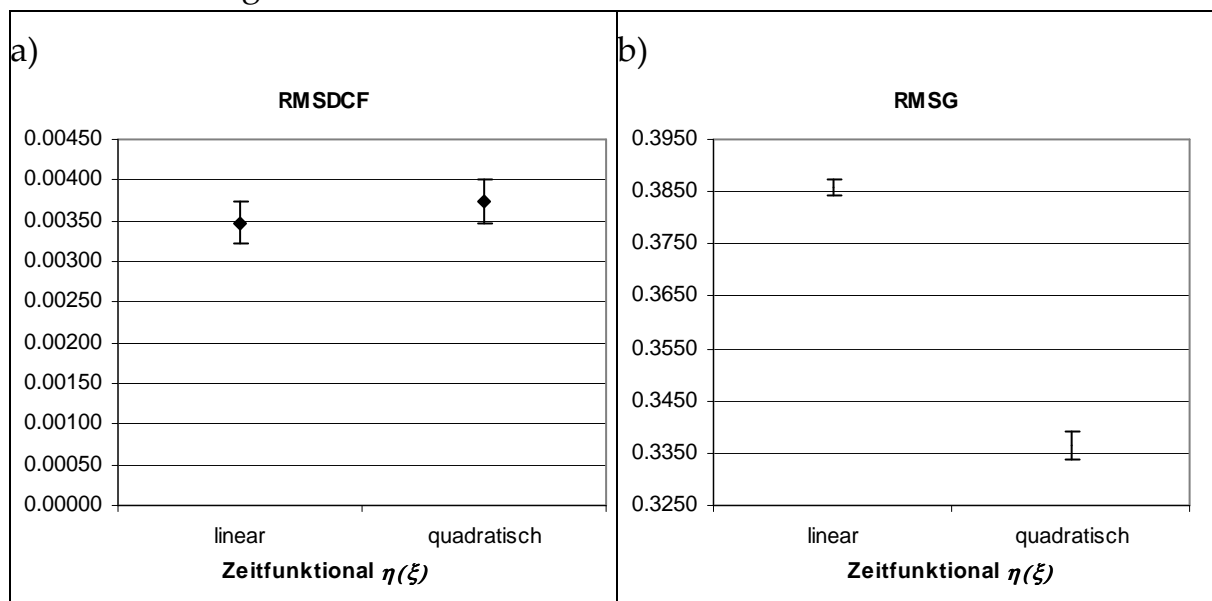
Abbildung 101: Beispielnetzarten Gridhyperfläche – Variation von  $\Delta_{\max}(1)$ , Heptan, links partiell optimiert unter Verwendung der Netzrestraints nach COLORIX1, rechts vollständig optimiert nach COLORIX2

### 2.8.4.1.4 Variation von $\eta(\xi)$

Tabelle 45: Variation der Zeitabhängigkeit des Adaptionsfaktors,  $\eta(\xi)$  nach Gleichung (2.17)

$\eta(\xi)$ (2.17)	$\emptyset$ RMSDCF	$\sigma$	relerr%	min	max	ndata
a) linear	0.00348	0.00027	7.7	0.00317	0.00389	5
b) quadratisch	0.00374	0.00027	7.2	0.00339	0.00403	5
$\eta(\xi)$ (2.17)	$\emptyset$ RMSG	$\sigma$	relerr%	min	max	ndata
a) linear	0.38571	0.001469	0.4	0.38436	0.38791	5
b) quadratisch	0.33649	0.002797	0.8	0.33305	0.33927	5
$\eta(\xi)$ (2.17)	$\emptyset$ RMSDSN	$\sigma$	relerr%	min	max	ndata
a) linear	0.12354	0.000099	0.1	0.12341	0.12366	5
b) quadratisch	0.13165	0.000191	0.1	0.13139	0.13193	5

Auch für den Adaptionsfaktors  $\eta(\xi)$  bestätigt sich in Abbildung 102 der Trend der schon beim Hexan zu erkennen war. RMSG- und RMSDSN-Werte zeigen einen gegenläufigen Einfluss des variierten Faktors, wobei der RMSDCF-Wert sich trendmäßig wie der RMSDSN-Wert verhält.



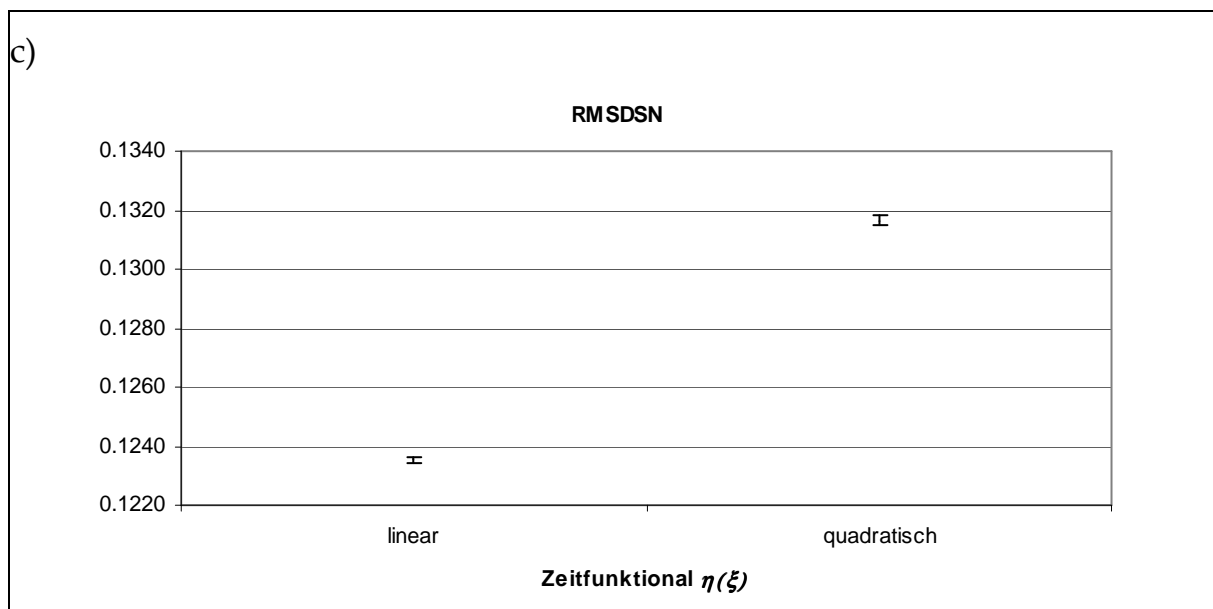


Abbildung 102: Variation von  $\eta(\xi)$  nach Gleichung (2.17)

Um zu ermitteln, welche der Netzkarten leichter zu interpretieren ist, werden in Abbildung 103 je ein Vertreter für beide Adaptionisfaktoren („linear“ und „quadratisch“) den MINIX1- und MINIX2-Modulen zugeführt. Insgesamt wirkt die „quadratische“ Netzkarte konsolidierter und erfüllt damit die Aufgabe der Datenreduktion besser.

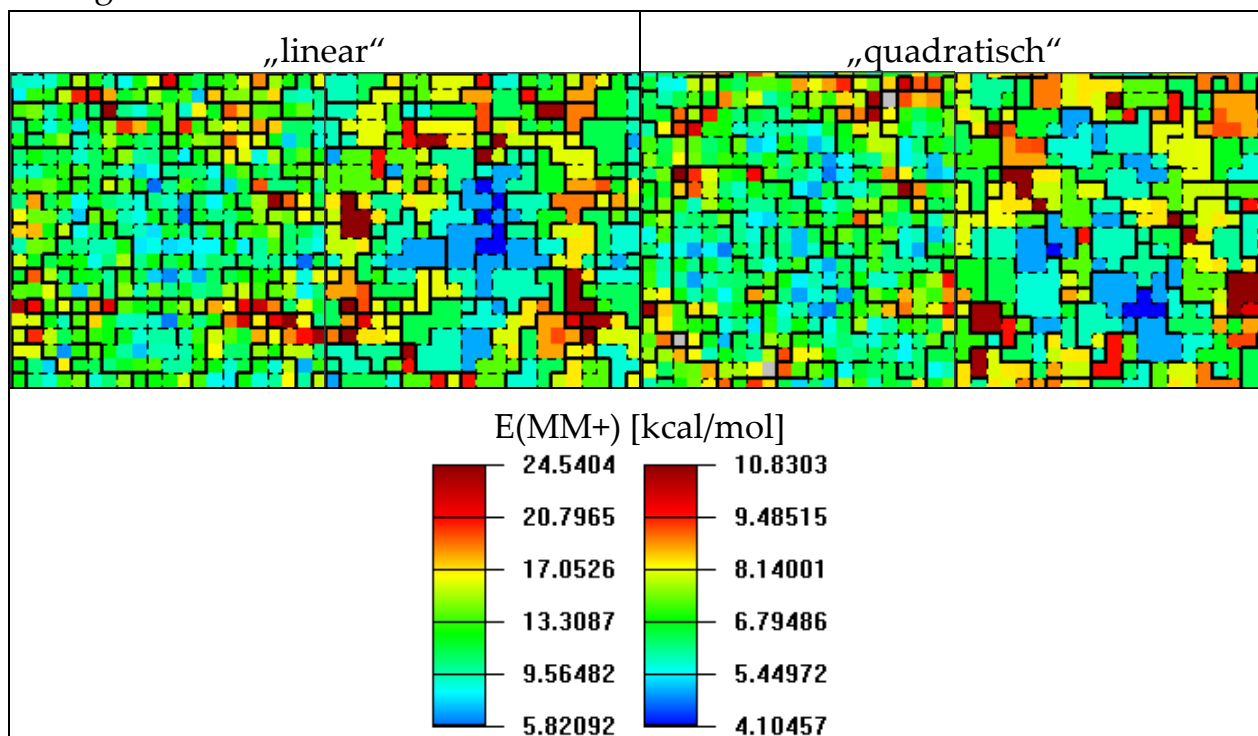


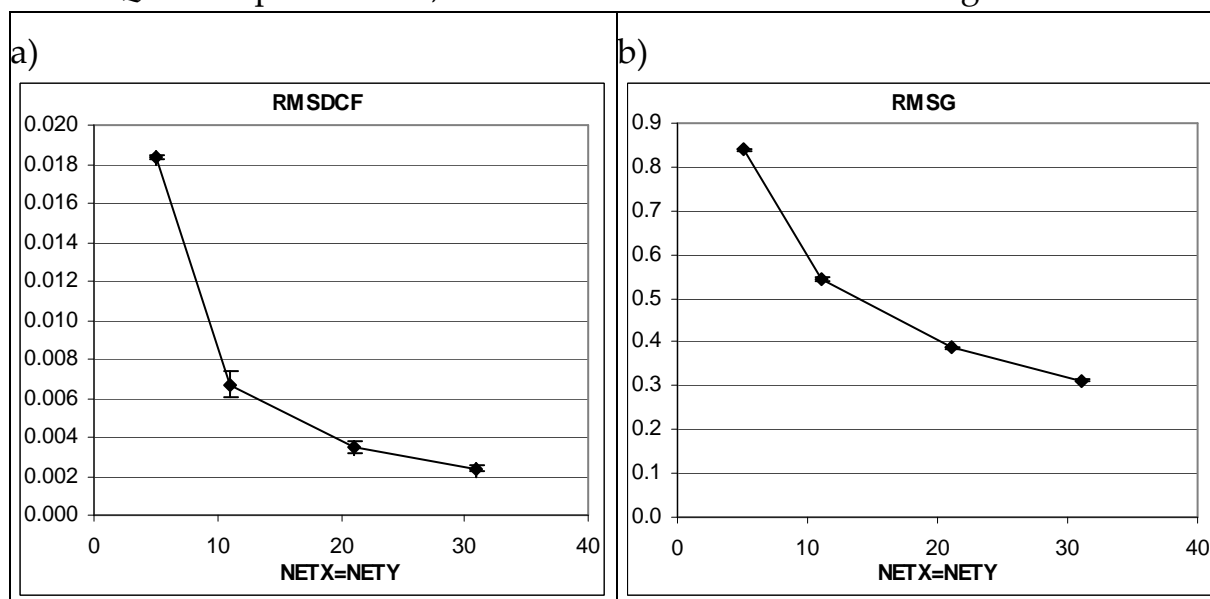
Abbildung 103: Beispielnetskarten Gridhyperfläche - Variation von  $\eta(\xi)$ , Heptan links partiell optimiert unter Verwendung der Netzrestraints nach COLORIX1, rechts vollständig optimiert nach COLORIX2

### 2.8.4.1.5 Variation von NETX, NETY

Tabelle 46: Variation der Netzgröße NETX=NETY

NETX=NETY	Ø RMSDCF	$\sigma$	relerr%	min	max	ndata
31	0.00239	0.00018	7.47	0.00212	0.00259	5
21	0.00350	0.00029	8.29	0.00306	0.00375	5
11	0.00670	0.00065	9.64	0.00608	0.00758	5
5	0.01838	0.00010	0.56	0.01822	0.01848	5
NETX=NETY	Ø RMSG	$\sigma$	relerr%	min	max	ndata
31	0.31268	0.00045	0.14	0.31212	0.31330	5
21	0.38618	0.00178	0.46	0.38428	0.38851	5
11	0.54349	0.00279	0.51	0.54066	0.54772	5
5	0.83849	0.00401	0.48	0.83281	0.84415	5
NETX=NETY	Ø RMSDSN	$\sigma$	relerr%	min	max	ndata
31	0.10130	0.00005	0.05	0.10122	0.10134	5
21	0.12348	0.00006	0.05	0.12341	0.12355	5
11	0.17063	0.00004	0.02	0.17059	0.17068	5
5	0.25126	0.00019	0.08	0.25098	0.25140	5

Wie bereits zuvor führen größere Netzkarten wenig überraschend zu verbesserten Qualitätsparametern, als Grafik ersichtlich in Abbildung 104.



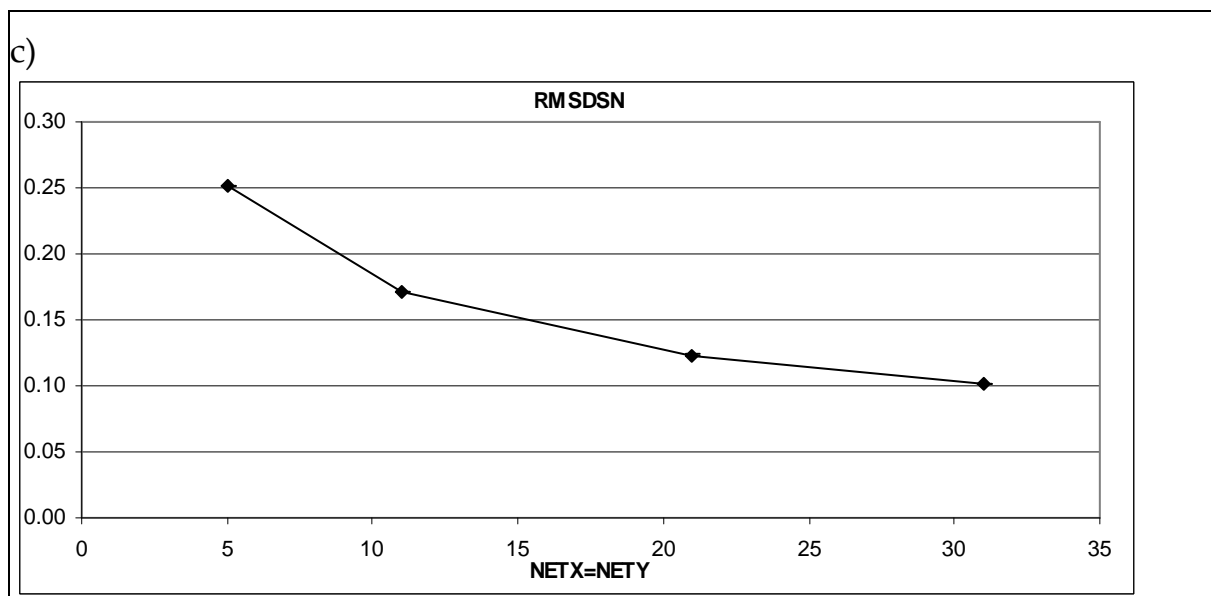


Abbildung 104: Variation der Netzgröße NETX=NETY

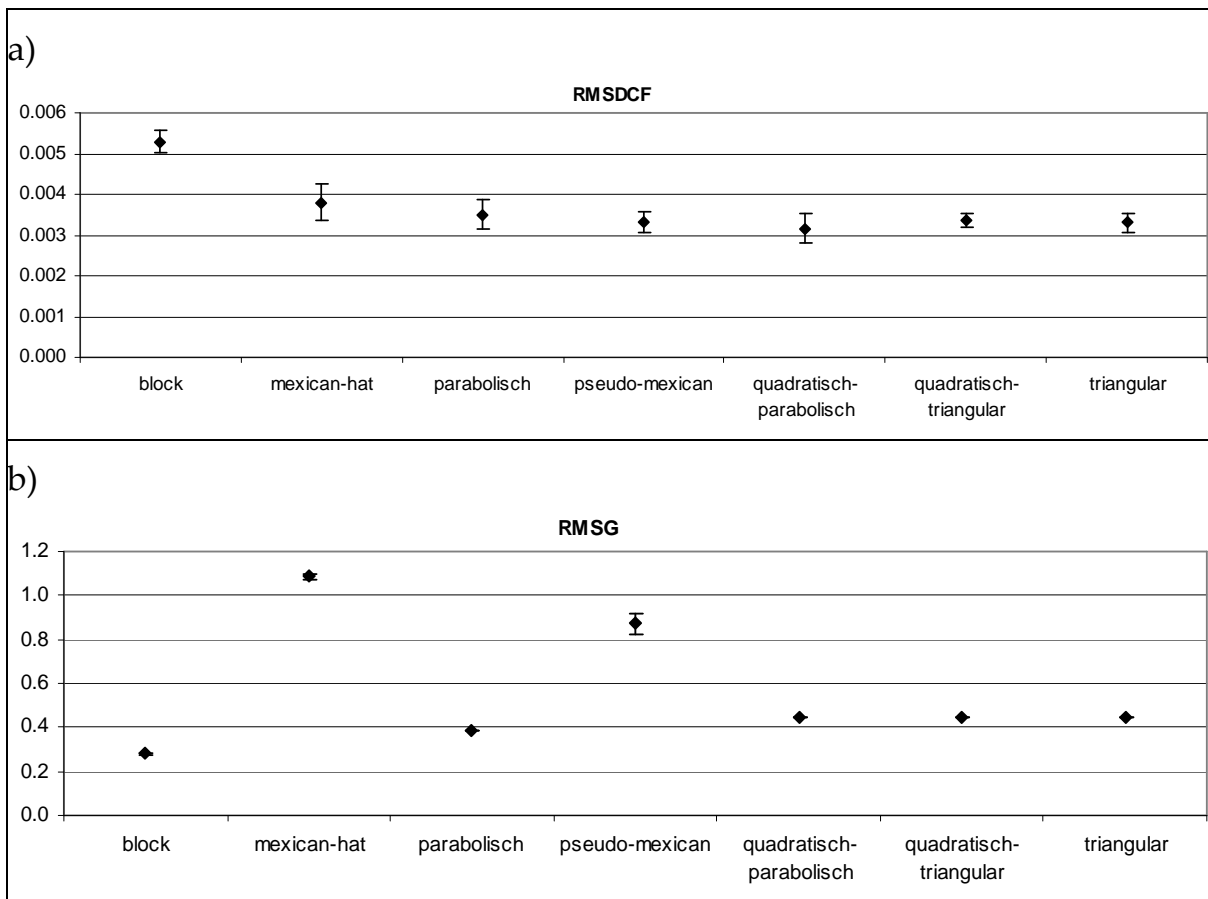
### 2.8.4.1.6 Variation der Abstandsfunktion $\vartheta(\zeta)$

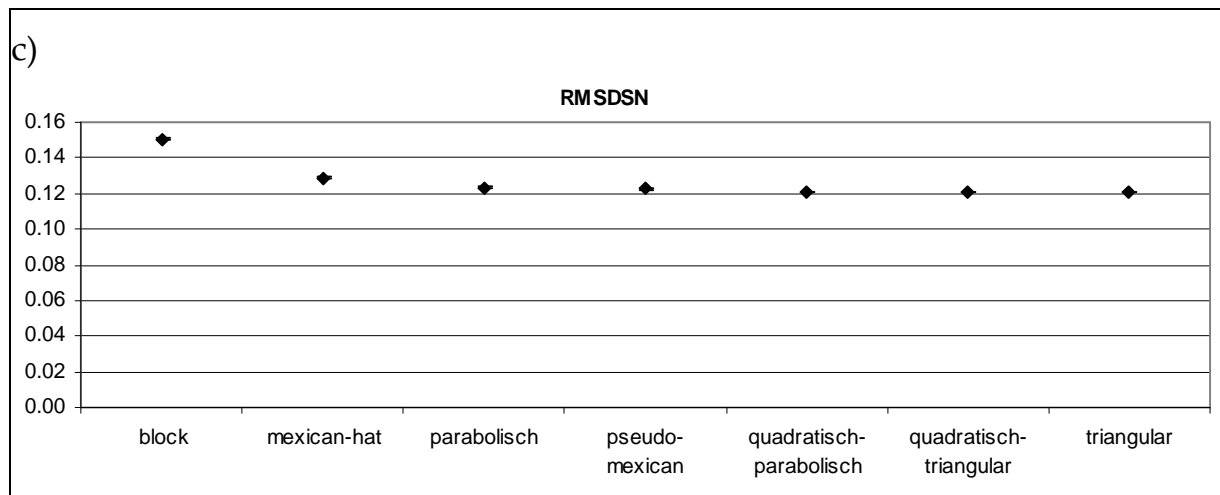
Tabelle 47: Variation der Abstandsfunktion  $\vartheta(\zeta)$  nach Gleichung (2.18)

$\vartheta(\zeta)$	$\bar{\varnothing}$ RMSDCF	$\sigma$	relerr%	min	max	ndata
<i>block</i>	0.00529	0.00028	5.25	0.00501	0.00566	5
<i>Mexican hat</i>	0.00379	0.00045	11.88	0.00313	0.00439	5
parabolisch	0.00350	0.00036	10.14	0.00293	0.00382	5
<i>pseudo Mexican</i>	0.00333	0.00026	7.86	0.00301	0.00361	5
quadratisch- parabolisch	0.00317	0.00036	11.23	0.00290	0.00379	5
quadratisch- triangular	0.00337	0.00018	5.28	0.00315	0.00357	5
triangular	0.00330	0.00022	6.63	0.00298	0.00356	5
$\vartheta(\zeta)$	$\bar{\varnothing}$ RMSG	$\sigma$	relerr%	min	max	ndata
<i>block</i>	0.27861	0.00118	0.42	0.27757	0.28054	5
<i>Mexican hat</i>	1.08593	0.01490	1.37	1.06657	1.10438	5
parabolisch	0.38490	0.00105	0.27	0.38318	0.38567	5
<i>pseudo Mexican</i>	0.87129	0.04622	5.30	0.79253	0.91333	5
quadratisch- parabolisch	0.44624	0.00270	0.60	0.44213	0.44931	5
quadratisch- triangular	0.44498	0.00308	0.69	0.44113	0.44906	5
triangular	0.44664	0.00296	0.66	0.44409	0.45168	5

$\mathcal{S}(\zeta)$	$\emptyset$ RMSDSN	$\sigma$	relerr%	min	max	ndata
<i>block</i>	0.15051	0.00060	0.40	0.14993	0.15144	5
<i>Mexican hat</i>	0.12869	0.00036	0.28	0.12811	0.12902	5
parabolisch	0.12353	0.00008	0.06	0.12346	0.12364	5
<i>pseudo Mexican</i>	0.12256	0.00024	0.20	0.12220	0.12283	5
quadratisch- parabolisch	0.12105	0.00009	0.07	0.12095	0.12116	5
quadratisch- triangular	0.12097	0.00010	0.08	0.12083	0.12111	5
triangular	0.12101	0.00006	0.05	0.12092	0.12107	5

Die Ergebnisse sind wiederum vergleichbar mit den Ergebnissen beim Hexan. „*Mexican hat*“-Funktion und „*pseudo Mexican*“ scheiden aufgrund von Fehlprojektionen aus. Die „*block*“-Funktion zeigt einen besonders niedrigen RMSG- und hohen RMSDSN- und RMSDCF-Wert.





**Abbildung 105: Variation der Abstandsfunktion  $v(\xi)$**

Beispielhafte Netzkarten nach MINIX1- und MINIX2-Modul sind in Abbildung 106 zu finden. Subjektiv am einfachsten auszuwerten sind die *block* Funktion und die parabolische Funktion, da es auf diesen Netzoberflächen die größeren und weniger strukturierten Cluster gibt. Dieser Trend ist deutlicher zu erkennen als noch beim Hexan.

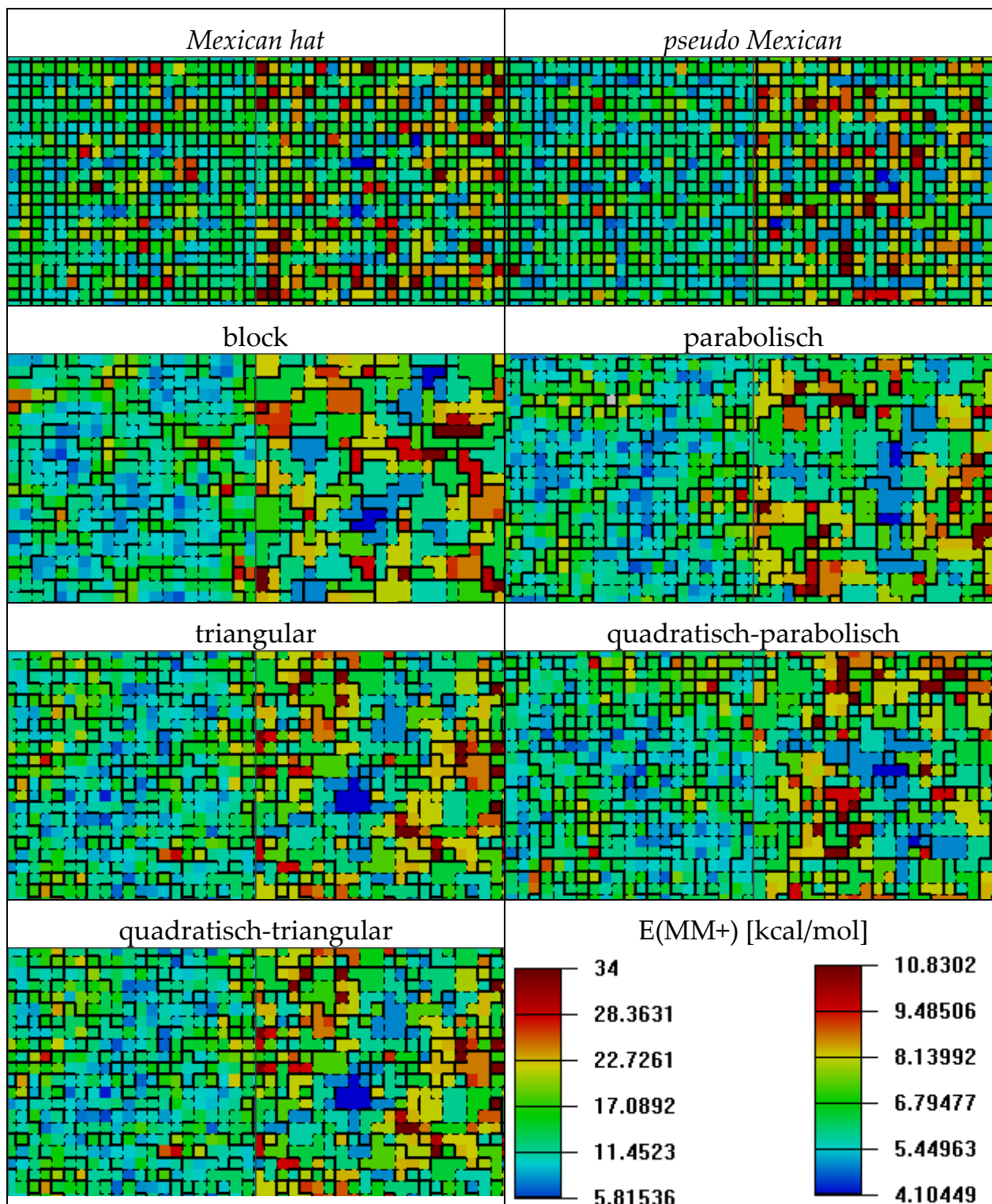


Abbildung 106: Beispielnetzarten Gridhyperfläche – Variation der Abstandsfunktion, Heptan links partiell optimiert unter Verwendung der Netzrestraints nach COLORIX1, rechts vollständig optimiert nach COLORIX2

#### 2.8.4.1.7 Fazit

Die Ergebnisse decken sich weitgehend mit den Ergebnissen, die beim Hexan erhalten wurden, so dass auf eine Wiederholung verzichtet wird. Es wurden folgende MTS Werte (minimale Trainingsschritte nach Kapitel 2.5.4) ermittelt:



- MTS(RMSG, 1%) = 10368000
- MTS(RMSDSN, 1%) = 6220800

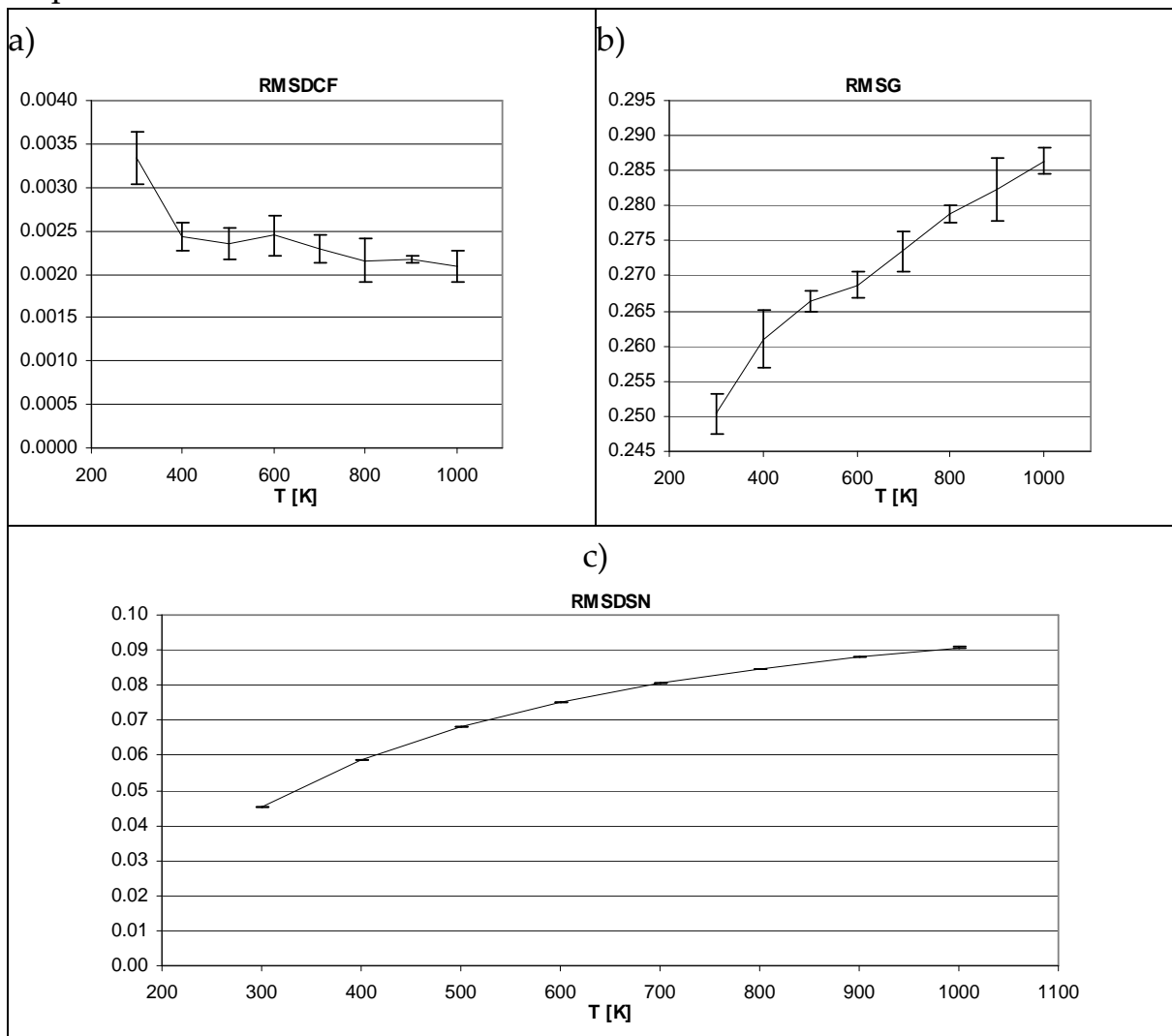
#### 2.8.4.2 MD Heptan, Training der Diederwinkel

Für das Heptan-Molekül wird bei unterschiedlichen Temperaturen eine Gruppe Moleküldynamikrechnungen durchgeführt. (Kraftfeld HYPERCHEM MM+, in vacuo, 200000 Strukturen, Simulationsdauer 100000 ps, Aufruf von FIX\_ROTATION.EXE bei je 10 ps, *step size* 0.001 ps, *datafreq*=1/500 *data steps*<sup>-1</sup>). Mit diesen MD-Eingabedaten werden neuronale Netze trainiert. [25x25 Neuronen; 100 *epochs*;  $\Delta_{\max}(1)=12$ ;  $f_0=0.4$ ; Abstandsfunktion parabolisch, Training der vier CC-Diederwinkel]. Die erhaltenen Qualitätsparameter sind in Tabelle 48 angeführt. Eine Netzkarte (700 K) mit namentlich identifizierten und den Clustern der Oberfläche zugeordneten Konformationen ist bereits in Abbildung 96 angeführt.

**Tabelle 48:** Heptan MD-Simulationen bei unterschiedlichen Temperaturen, Training der zentralen Diederwinkel

T [K]	Ø RMSDCF	$\sigma$	relerr%	min	max	ndata
1000	0.00209	0.00017	8.35	0.00181	0.00225	5
900	0.00216	0.00004	1.74	0.00211	0.00221	5
800	0.00216	0.00025	11.74	0.00188	0.00249	5
700	0.00229	0.00017	7.31	0.00206	0.00254	5
600	0.00245	0.00023	9.56	0.00218	0.00271	5
500	0.00235	0.00019	7.95	0.00217	0.00266	5
400	0.00244	0.00016	6.64	0.00221	0.00266	5
300	0.00333	0.00030	8.98	0.00285	0.00361	5
T [K]	Ø RMSG	$\sigma$	relerr%	min	max	ndata
1000	0.28640	0.00194	0.68	0.28373	0.28829	5
900	0.28235	0.00453	1.60	0.27670	0.28892	5
800	0.27889	0.00129	0.46	0.27748	0.28100	5
700	0.27360	0.00286	1.05	0.27048	0.27744	5
600	0.26867	0.00184	0.68	0.26732	0.27189	5
500	0.26638	0.00143	0.54	0.26509	0.26854	5
400	0.26102	0.00405	1.55	0.25637	0.26717	5
300	0.25040	0.00279	1.12	0.24846	0.25528	5
T [K]	Ø RMSDSN	$\sigma$	relerr%	min	max	ndata
1000	0.09079	0.00002	0.03	0.09076	0.09082	5
900	0.08800	0.00004	0.05	0.08795	0.08805	5
800	0.08461	0.00008	0.09	0.08450	0.08470	5
700	0.08046	0.00003	0.04	0.08042	0.08050	5
600	0.07509	0.00004	0.06	0.07502	0.07512	5
500	0.06832	0.00004	0.06	0.06827	0.06839	5
400	0.05881	0.00009	0.15	0.05870	0.05894	5
300	0.04521	0.00012	0.27	0.04506	0.04539	5

Die grafische Darstellung in Abbildung 107 zeigt wiederum steigende RMSG- und RMSDSN-Werte bei höheren Temperaturen. Dies lässt sich erklären durch den insgesamt mannigfaltigeren Konformationsraum in den Eingabedaten bei höheren Temperaturen, sodass die Netzoberfläche gemessen in der Zahl der Neuronen relativ gesehen kleiner wird. Entsprechend verhalten sich die RMSG- und RMSDSN-Graphen analog wie bei der Variation der Netzgröße in Kapitel 2.8.4.1.5.



**Abbildung 107: Variation der Temperatur – Heptan MD, Training der drei CC-Diederwinkel**

Der sinkende RMSDCF-Wert kann durch Abbildung 108 erklärt werden. Im Gegensatz zu den kleineren Molekülen enthalten bereits die Simulationen bei niedrigen Temperaturen praktisch alle Konformationen zu kleinen Anteilen. Auf der Netzoberfläche bewirkt dies, dass relativ kleine Bereiche der Hyperfläche überproportional stark repräsentiert werden, was zu einem hohen RMSDCF-Wert bei niedrigen Temperaturen führt. Bei höheren Temperaturen ist die Aufenthaltswahrscheinlichkeit des Moleküls gleichmäßiger über die energetisch unterschiedlichen Konformationen verteilt, was eine Projektion mit niedrigerem RMSDCF-Wert zulässt.

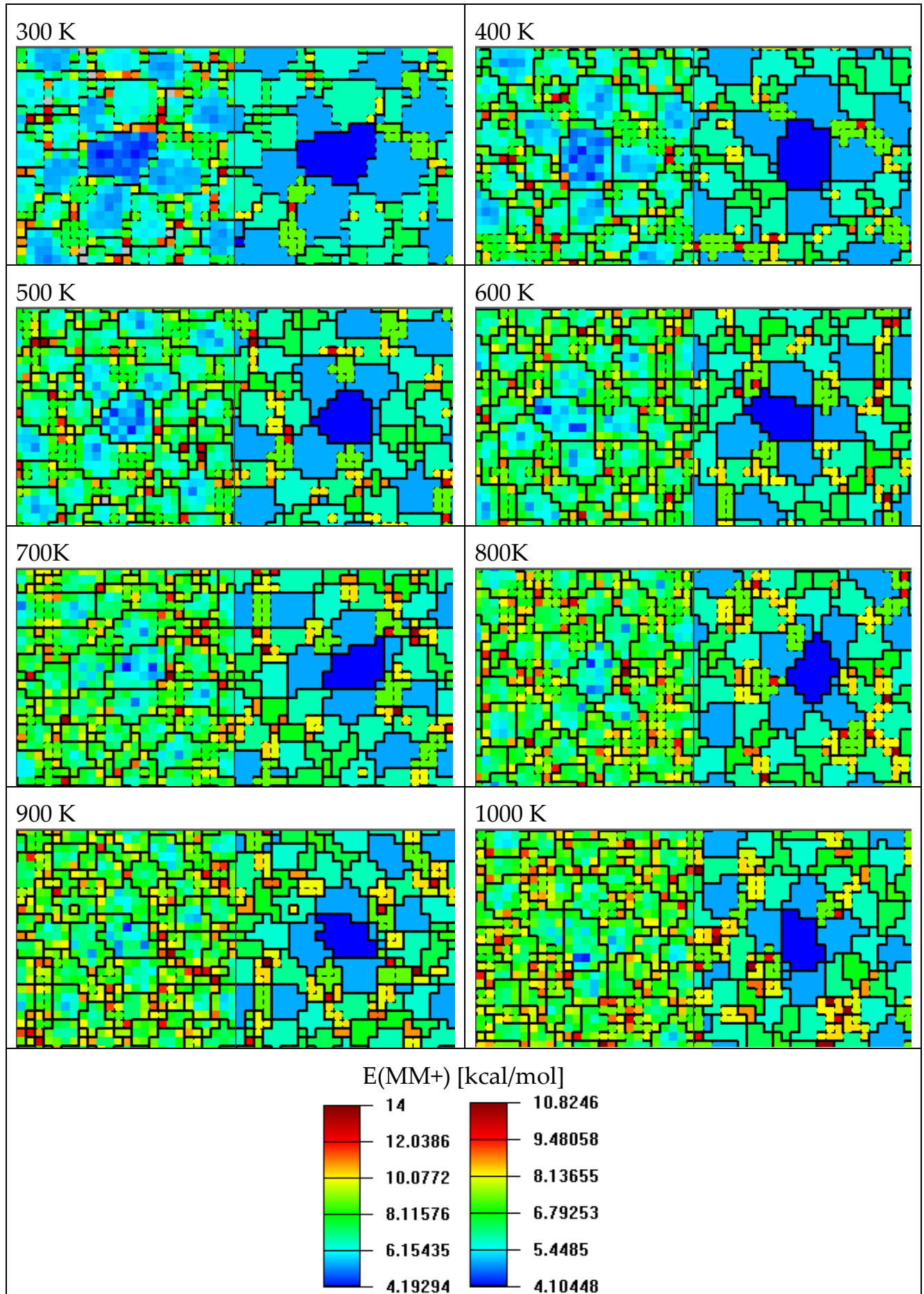


Abbildung 108: Beispielnetskarten Gridhyperfläche – Variation der Temperatur – Heptan MD, Training der CC-Diederwinkel, links partiell optimiert nach COLORIX1, rechts vollständig optimiert nach COLORIX2

### 2.8.4.3 MD Heptan, Training 42 zufälliger Distanzen

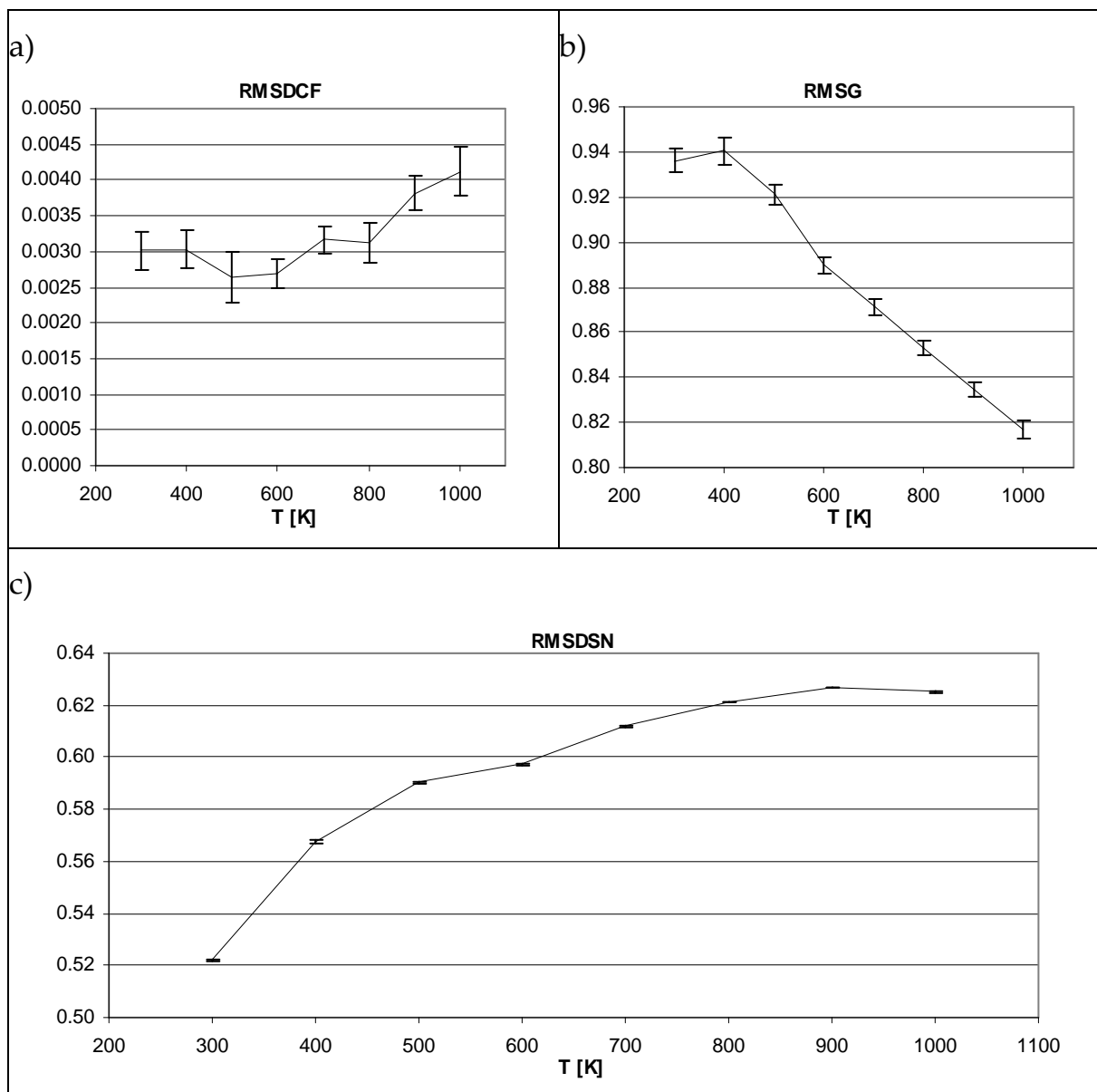
Repräsentativ für die komplette Abstandsmatrix aller C-C, C-H und H-H Abstände wird beim Heptan aus Rechenzeitgründen ein Satz von zweiundvierzig zufällig ausgewählten C-C, C-H und H-H Abständen (die doppelte Anzahl aller Schweratomsabstände) trainiert. Als Eingabedaten werden die gleichen MD-Rechnungen wie im vorherigen Kapitel verwendet.

Die statistischen Daten des Netztrainings sind in Tabelle 49 aufgeführt. [25x25 Neuronen; 20 *epochs*;  $\Delta_{\max}(1)=12$ ;  $f_0=0.4$ ; Abstandsfunktion parabolisch, Training 42 zufällig ausgewählter Distanzen].

**Tabelle 49: Heptan MD-Simulationen bei unterschiedlichen Temperaturen, Training 42 zufällig ausgewählter Distanzen**

T [K]	Ø RMSDCF	$\sigma$	relerr%	min	max	ndata
1000	0.00412	0.00034	8.30	0.00378	0.00457	5
900	0.00382	0.00025	6.56	0.00363	0.00425	5
800	0.00313	0.00028	9.01	0.00282	0.00355	5
700	0.00317	0.00019	5.98	0.00296	0.00344	5
600	0.00269	0.00021	7.84	0.00256	0.00305	5
500	0.00264	0.00036	13.79	0.00229	0.00319	5
400	0.00303	0.00026	8.71	0.00281	0.00345	5
300	0.00301	0.00026	8.76	0.00272	0.00340	5
T [K]	Ø RMSG	$\sigma$	relerr%	min	max	ndata
1000	0.81677	0.00409	0.50	0.81279	0.82235	5
900	0.83479	0.00329	0.39	0.82950	0.83860	5
800	0.85322	0.00321	0.38	0.85028	0.85836	5
700	0.87135	0.00362	0.42	0.86699	0.87612	5
600	0.88973	0.00380	0.43	0.88521	0.89556	5
500	0.92121	0.00438	0.48	0.91385	0.92473	5
400	0.94034	0.00571	0.61	0.93291	0.94865	5
300	0.93612	0.00526	0.56	0.92857	0.94135	5
T [K]	Ø RMSDSN	$\sigma$	relerr%	min	max	ndata
1000	0.62530	0.00033	0.05	0.62503	0.62585	5
900	0.62675	0.00018	0.03	0.62655	0.62697	5
800	0.62107	0.00017	0.03	0.62084	0.62124	5
700	0.61199	0.00024	0.04	0.61161	0.61225	5
600	0.59727	0.00032	0.05	0.59703	0.59783	5
500	0.59054	0.00034	0.06	0.59012	0.59105	5
400	0.56741	0.00056	0.10	0.56698	0.56836	5
300	0.52211	0.00035	0.07	0.52173	0.52254	5

Der RMSDSN-Wert steigt mit der Temperatur, während der RMSG-Wert bei 400K ein Maximum erreicht und dann kontinuierlich absinkt, wie in Abbildung 91 zu sehen ist.



**Abbildung 109: Variation der Temperatur – Heptan MD, Training 42 zufällig ausgewählter Distanzen**

Optimierte Beispielnetskarten nach COLORIX1/COLORIX2 sind in Abbildung 110 dargestellt. Die *all-anti* Konformation ist meist doppelt auf der Netzoberfläche vertreten, ein Phänomen, das bereits von den Netskarten aus Kapitel 2.8.3.2 bekannt ist. Interessanter Ausreißer ist die Alternativtopologie der 500 K-Netskarte, welche die sterischen Probleme bei der Anordnung der hellblauen *single-gauche* Konformationen um einen einzelnen *all-anti* Cluster demonstriert.

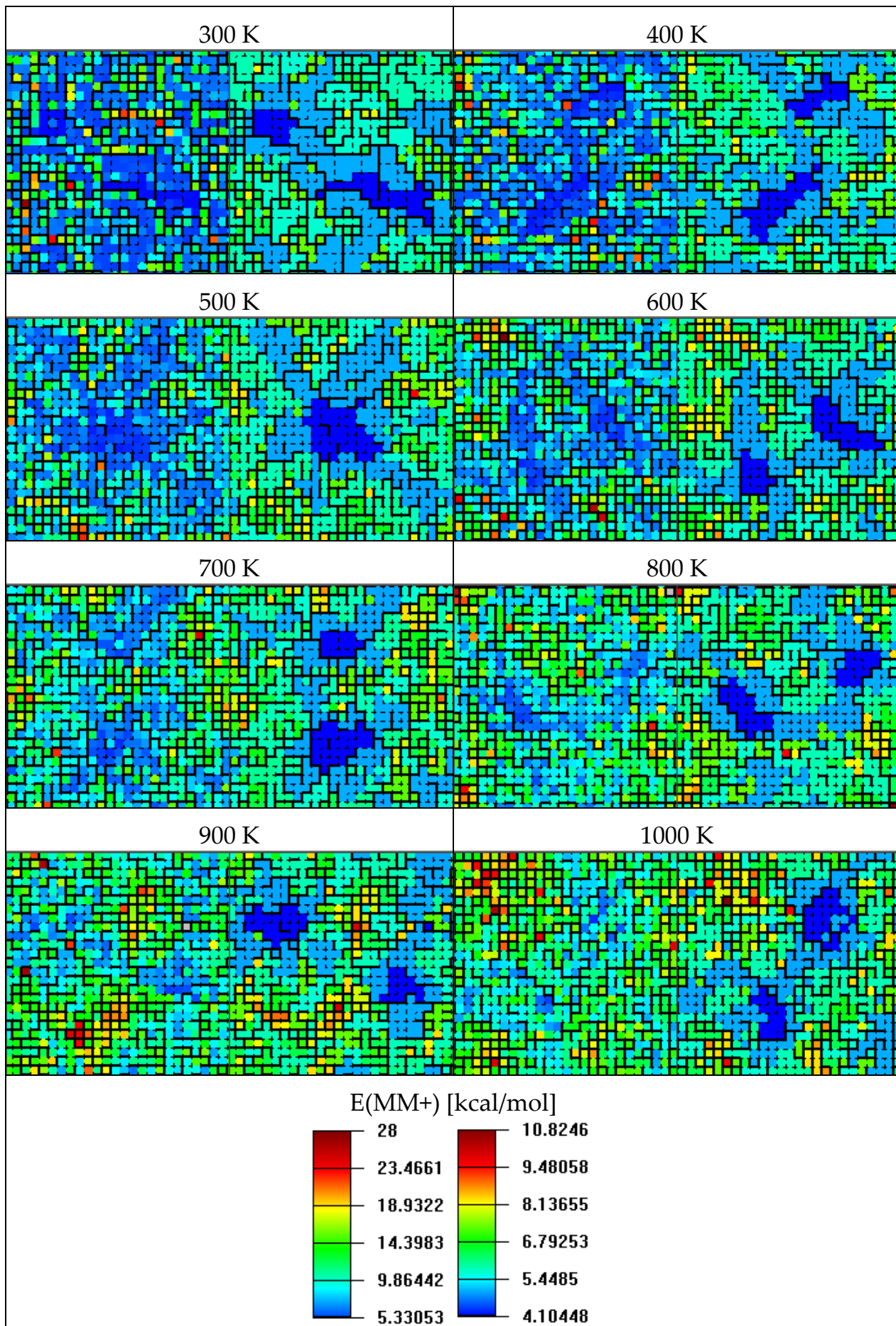


Abbildung 110: Beispielnetskarten Gridhyperfläche – Variation der Temperatur – Heptan MD, Training 42 zufällig ausgewählter Distanzen, links partiell optimiert nach COLORIX1, rechts frei optimiert nach COLORIX2

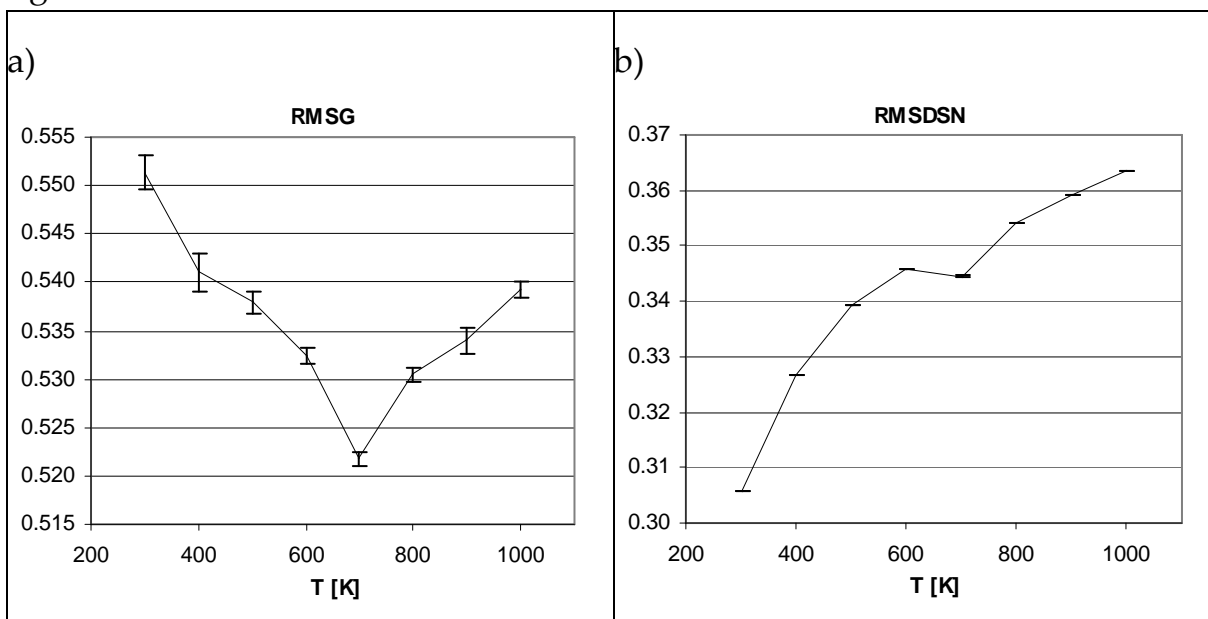
### 2.8.4.4 MD Heptan, Training der Schweratomdistanzen

Das Training der Schweratommatrix führt wiederum zu einer vereinfachten Hyperfläche. Die statistischen Daten des Netztrainings [25x25 Neuronen; 100 epochs;  $\Delta_{\max}(1)=12$ ;  $f_0=0.4$ ; Abstandsfunktion parabolisch, Training der CC-Distanzen] sind in Tabelle 50 aufgeführt.

**Tabelle 50: Heptan MD-Simulationen bei unterschiedlichen Temperaturen, Training der Schweratomdistanzmatrix**

T [K]	Ø RMSG	$\sigma$	relerr%	min	max	ndata
1000	0.50237	0.00188	0.37	0.50072	0.50523	5
900	0.49704	0.00083	0.17	0.49569	0.49779	5
800	0.48872	0.00103	0.21	0.48739	0.48983	5
700	0.49560	0.00102	0.21	0.49427	0.49681	5
600	0.48522	0.00155	0.32	0.48346	0.48713	5
500	0.47859	0.00088	0.18	0.47745	0.47950	5
400	0.47771	0.00135	0.28	0.47602	0.47933	5
300	0.43714	0.00125	0.29	0.43589	0.43882	5
T [K]	Ø RMSDSN	$\sigma$	relerr%	min	max	ndata
1000	0.29602	0.00009	0.03	0.29589	0.29611	5
900	0.29225	0.00016	0.05	0.29207	0.29243	5
800	0.28589	0.00009	0.03	0.28581	0.28604	5
700	0.28739	0.00004	0.01	0.28734	0.28745	5
600	0.27803	0.00012	0.04	0.27791	0.27816	5
500	0.26662	0.00009	0.03	0.26649	0.26671	5
400	0.25655	0.00008	0.03	0.25646	0.25663	5
300	0.21018	0.00008	0.04	0.21011	0.21031	5

Die grafische Auftragung obiger Werte ist in Abbildung 111 zu finden. Auffällig ist ein scharfes Minimum der RMSG-Werte bei 700K.



**Abbildung 111: Variation der Temperatur – Heptan MD, Training der Schweratomdistanzmatrix**

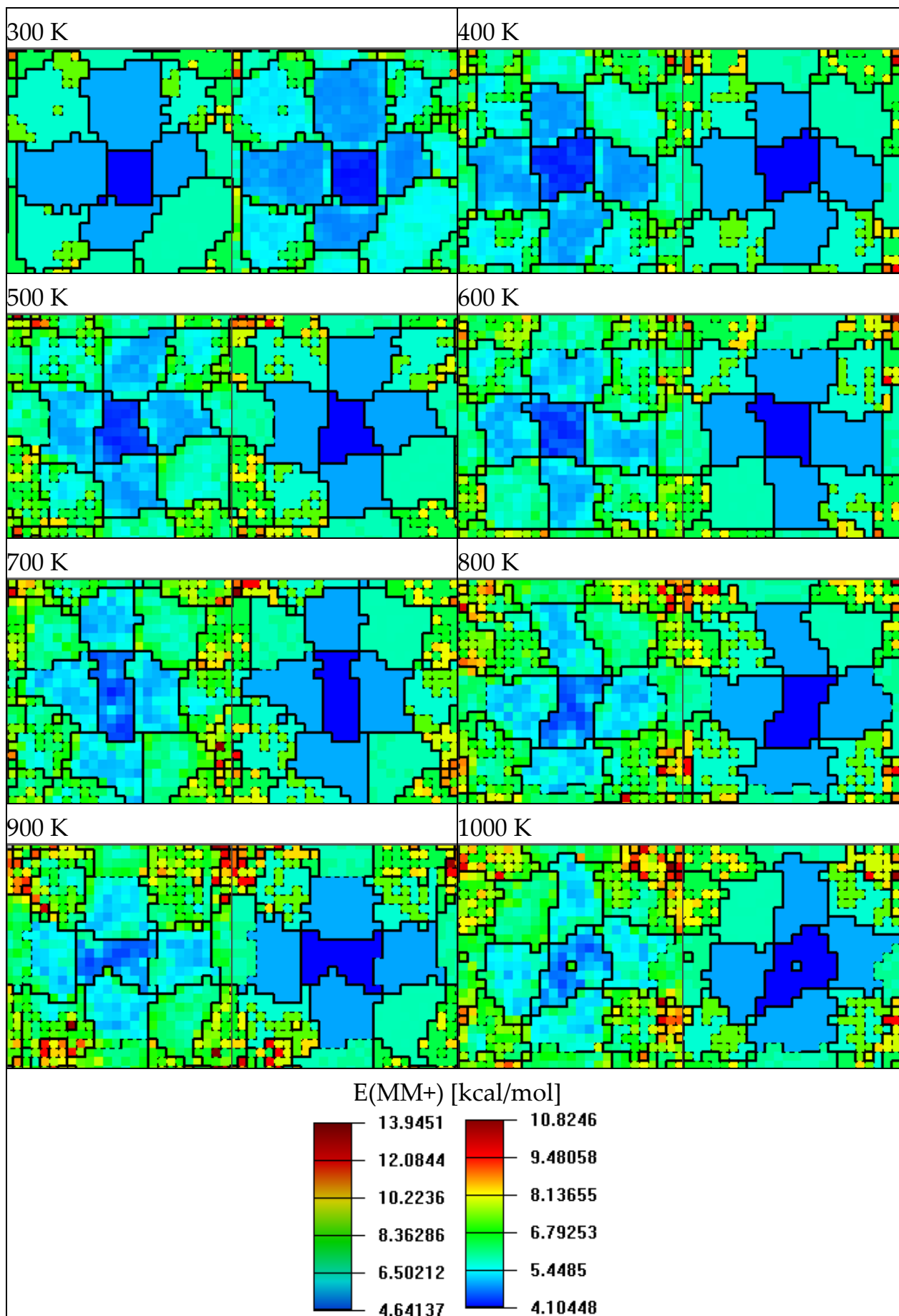


Abbildung 112: Beispielnetzarten Gridhyperfläche – Variation der Temperatur – Heptan MD, Training der Schweratomdistanzmatrix, links partiell optimiert unter Verwendung der Netzrestraints nach COLORIX1, rechts vollständig optimiert nach COLORIX2



Die optimierten Netzkarten in Abbildung 112 geben keinen Aufschluss über die Ursache für den Knick im RMSG-Graphen. Eine teilweise Zuordnung der Cluster der 700 K-Netzkarte ist bereits in Abbildung 96 aufgetragen. Die Anzahl der durch die CC-Abstandsmatrix nicht mehr unterscheidbaren Konformationen ist größer als bei den vorherigen Molekülen. In Abbildung 112 ist außerdem deutlich das Ansteigen der Clusterfläche der Hochenergiekonformationen bei höheren Temperaturen ersichtlich.

### 2.8.5 *n*-Oktan

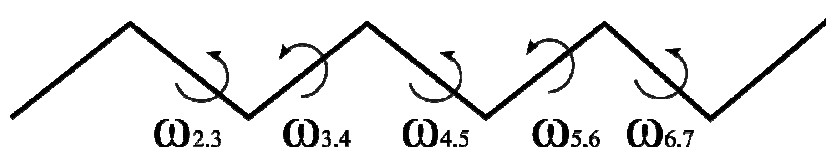


Abbildung 113: CC-Diederwinkel im *n*-Oktan

Mit fünf frei rotierbaren Diederwinkeln kann die Anzahl der Konformationen zu  $3^5=243$  abgeschätzt werden. Eine übersichtliche Darstellung des Konformationsraums ist auf dem Papier klassisch nicht mehr möglich.

Wie beim Heptan werden die Cluster in Abbildung 114 auf exemplarischen Netzkarten mit Hilfe des Skriptes PRINT\_SELECTION\_VALUES\_GUI.TK nach Konformationen klassifiziert.

In Abbildung 114 b) sind nach Training der CC-Abstände Enantiomere grundsätzlich nicht unterscheidbar, jeder Cluster mit z.B. gaaaa Konformationen repräsentiert selbstverständlich auch Gaaaa Strukturen, auch wenn dies nicht explizit in die Grafik eingetragen wird. Des Weiteren sind teilweise mehrere Konformere, welche sich nur durch Groß/Kleinschreibung des Buchstaben ‚G‘ ( $G=gauche^+$ ,  $g=gauche^-$ ) unterscheiden, in einem Cluster aufzufinden. Die Tatsache, dass diese auf der Netzkarte nicht immer alle aufgefunden werden können, heißt nicht zwingend, dass diese nicht vom entsprechenden Bereich repräsentiert werden.

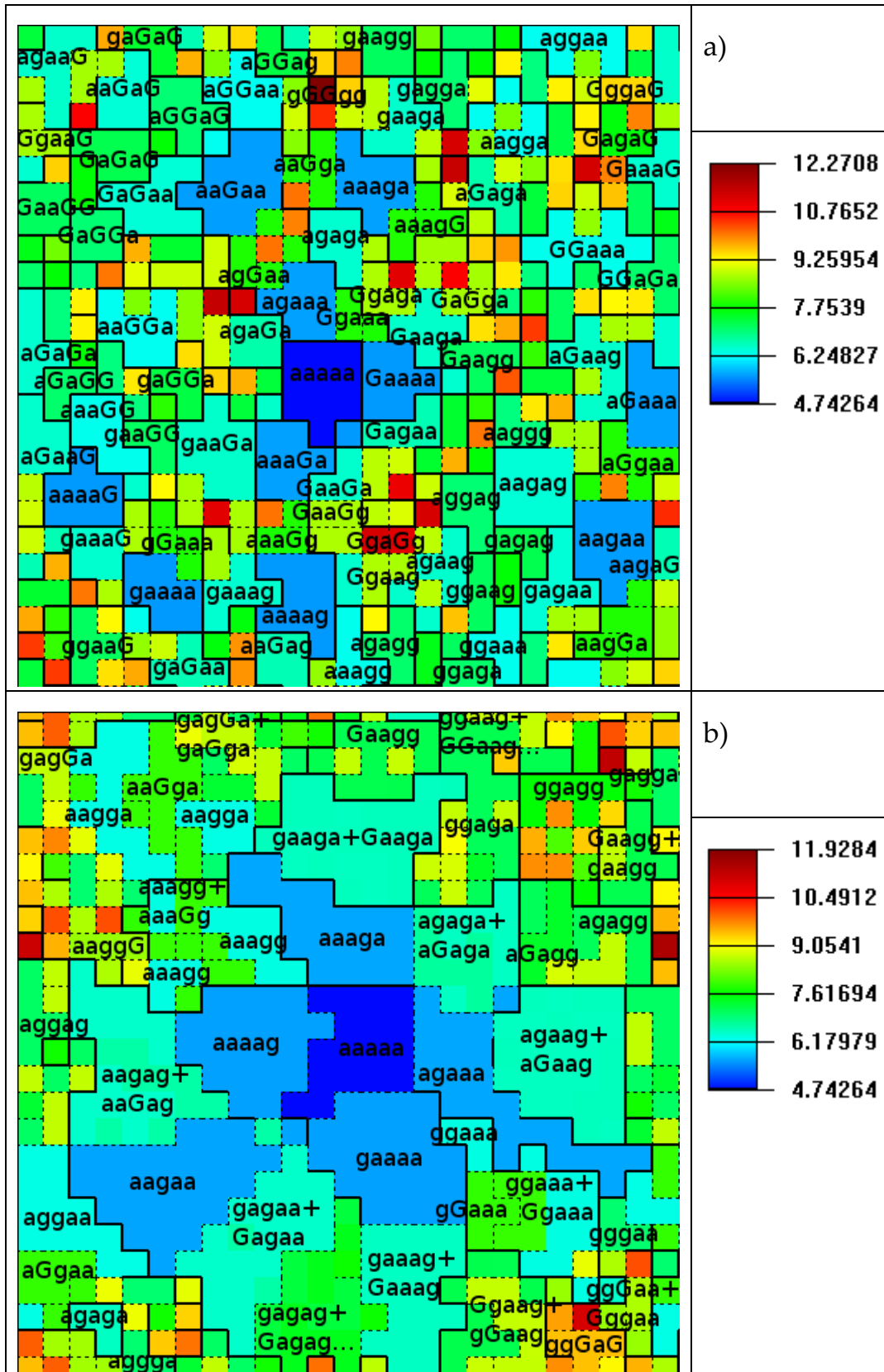


Abbildung 114: Beispielprojektion der MM+ Hyperfläche des Oktans und Energieskala in kcal/mol

a) Training der Diederwinkel, MD 800K

b) Training der Schweratommatrix, MD 800K

### 2.8.5.1 Training der Grid-Hyperfläche

Über die in Abbildung 113 angegebenen Diederwinkel wird mit dem Programm GRID.TCL (5.2.2) eine Gridrechnung in 60°-Schritten durchgeführt. [Kraftfeld: HYPERCHEM MM+]

- $\omega_{2,3}$ : 0, 60, 120, ..., 300
- $\omega_{3,4}$ : 0, 60, 120, ..., 300
- $\omega_{4,5}$ : 0, 60, 120, ..., 300
- $\omega_{5,6}$ : 0, 60, 120, ..., 300
- $\omega_{6,7}$ : 0, 60, 120, ..., 300

Die RMSDCF-Werte werden durch Klassifizierung der Oktankonformationen mittels des Programms CLASSIFY\_SNP\_CONFORMATIONS.EXE, siehe Kapitel 5.3.5, durchgeführt.

Es kommen wiederum die gleichen Vorgabewerte für das Training der Kohonennetze zum Einsatz. [ $f_0=0.15$ ,  $epochs=1000$ ,  $neighborhood\ function=parabolisch$ ,  $NETX=NETY=21$ ,  $linear\ learning\ decay$ ].

Die Netzoberflächen nach den MINIX1- und MINIX2-Modulen sind zu stark fragmentiert, um eine subjektive Beurteilung zuzulassen.

#### 2.8.5.1.1 Variation der epochs

Tabelle 51: Oktan - Variation epochs

epochs	Ø RMSDCF	$\sigma$	relerr%	min	max	ndata
1000	0.00266	0.00015	5.52	0.00245	0.00283	5
500	0.00257	0.00004	1.51	0.00251	0.00262	5
300	0.00274	0.00004	1.42	0.00271	0.00280	5
200	0.00299	0.00004	1.50	0.00292	0.00303	5
100	0.00326	0.00021	6.29	0.00294	0.00344	5
50	0.00337	0.00013	3.74	0.00322	0.00350	5
10	0.00396	0.00014	3.43	0.00379	0.00410	5
2	0.00503	0.00016	3.09	0.00487	0.00520	5
1	0.00584	0.00024	4.14	0.00559	0.00613	5

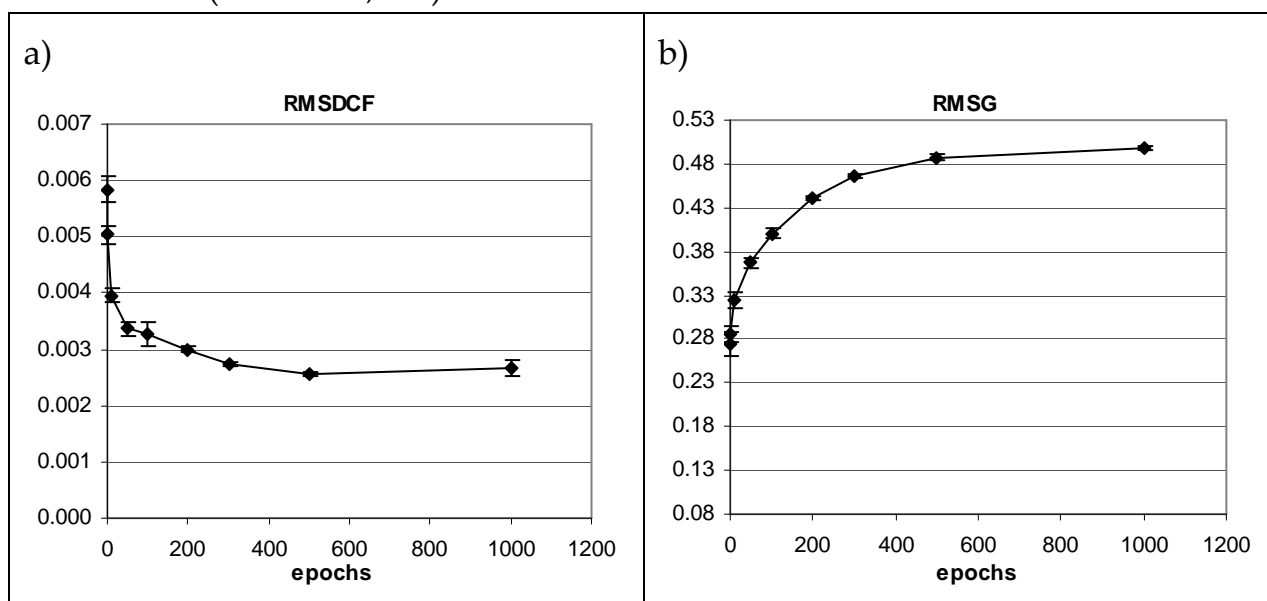
<i>epochs</i>	Ø RMSG	$\sigma$	relerr%	min	max	ndata
1000	0.49797	0.00185	0.37	0.49550	0.49985	5
500	0.48740	0.00371	0.76	0.48317	0.49170	5
300	0.46664	0.00208	0.45	0.46422	0.46927	5
200	0.44171	0.00229	0.52	0.43906	0.44504	5
100	0.40039	0.00545	1.36	0.39265	0.40787	5
50	0.36669	0.00558	1.52	0.35998	0.37273	5
10	0.32507	0.00868	2.67	0.31833	0.34019	5
2	0.28592	0.00942	3.30	0.27261	0.29923	5
1	0.27400	0.01324	4.83	0.25895	0.28759	5
<i>epochs</i>	Ø RMSDSN	$\sigma$	relerr%	min	max	ndata
1000	0.18329	0.00025	0.14	0.18310	0.18371	5
500	0.18538	0.00019	0.10	0.18509	0.18562	5
300	0.18865	0.00032	0.17	0.18838	0.18918	5
200	0.19249	0.00027	0.14	0.19227	0.19286	5
100	0.20163	0.00045	0.22	0.20100	0.20215	5
50	0.21098	0.00059	0.28	0.21007	0.21162	5
10	0.22992	0.00132	0.58	0.22799	0.23157	5
2	0.25756	0.00232	0.90	0.25402	0.26047	5
1	0.26917	0.00189	0.70	0.26665	0.27145	5

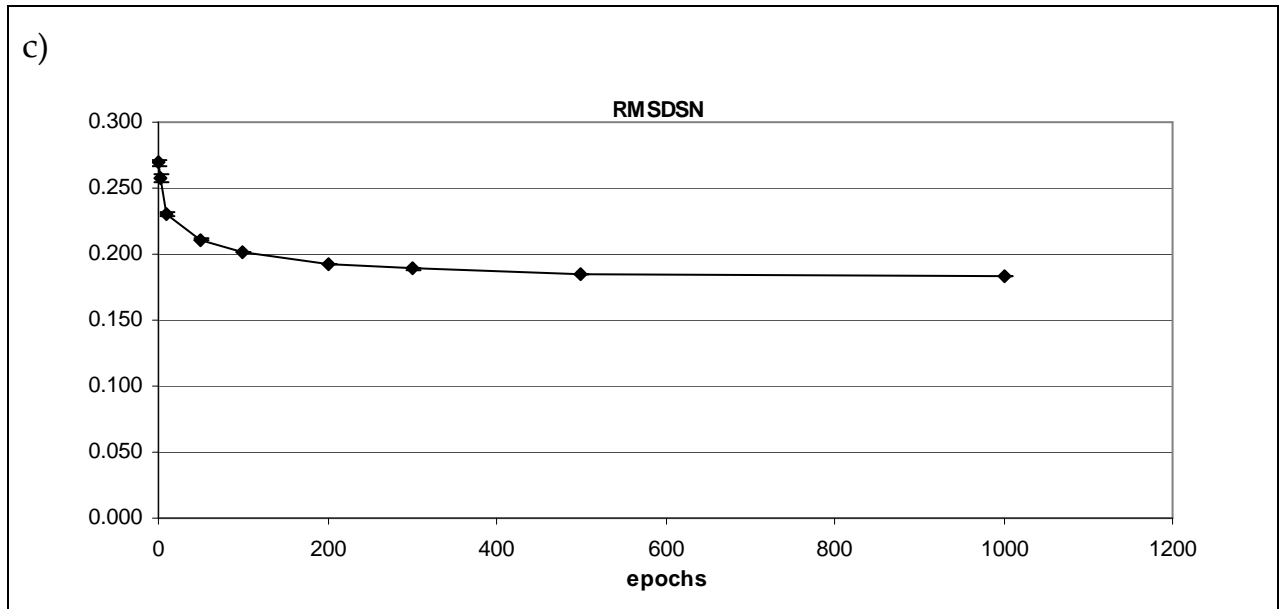
Die grafische Auftragung der obigen Tabellenwerte ergibt identische Trends wie bereits beim Heptan.  $\text{RMSG}_{\text{end}}^{[109]}$  und  $\text{RMSDSN}_{\text{end}}^{[110]}$  werden extrapoliert mit dem Programm ORIGIN<sup>[105]</sup> für alle *epochs* > 10. Dabei wird erkannt, dass bei keinem der Trainingsvorgänge genügend *epochs* trainiert werden, um das Netztraining zu „beenden“, sprich auf 1% an den finalen Wert heranzuführen.

$$\text{RMSG}_{\text{end}} = 0.51274 \pm 0.00587 \quad \text{epoch}_{\text{end}}(\text{RMSG}) > 1000$$

$$\text{RMSDSN}_{\text{end}} = 0.18029 \pm 0.00095 \quad \text{epoch}_{\text{end}}(\text{RMSG}) > 1000$$

- $\text{MTS}(\text{RMSG}, 1\%) > 7776000$
- $\text{MTS}(\text{RMSDSN}, 1\%) > 7776000$



Abbildung 115: Variation der *epochs*, a) RMSDCF, b) RMSG, c) RMSDSN

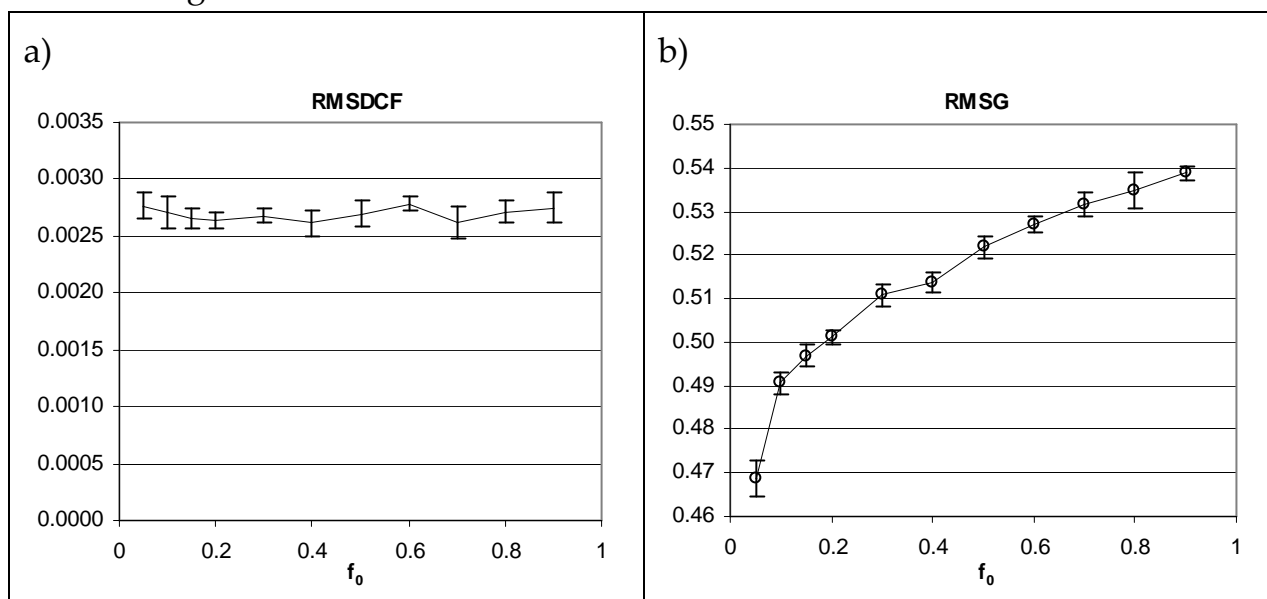
### 2.8.5.1.2 Variation von $f_0$

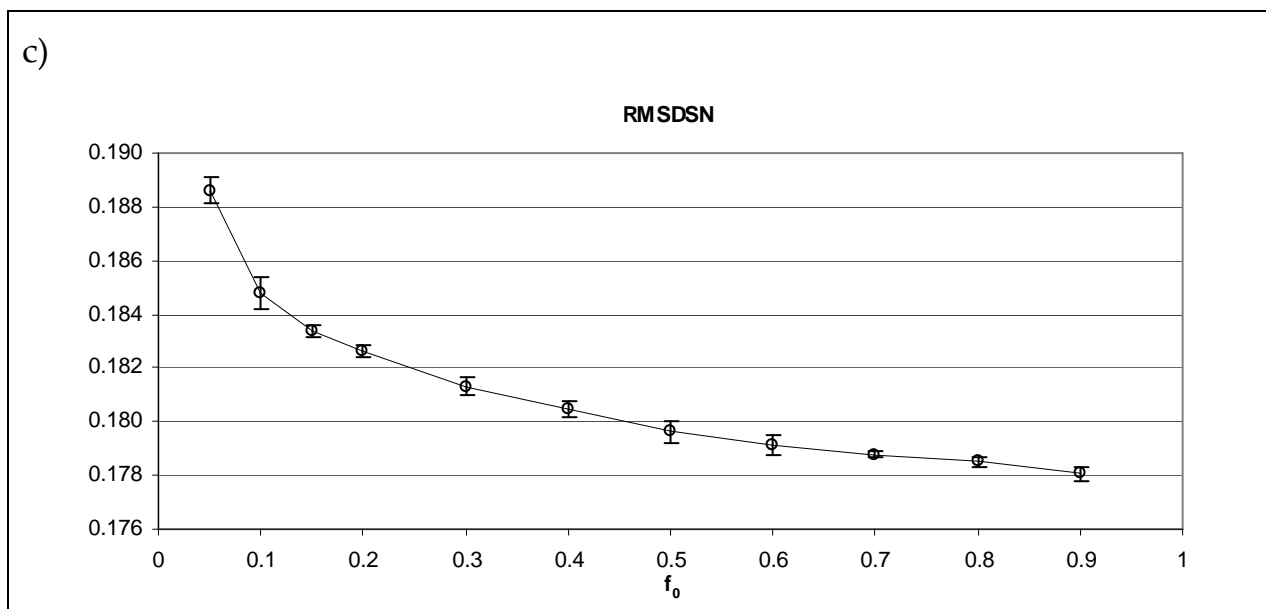
Tabelle 52: Variation des maximalen Adaptionfaktors  $f_0$  nach Gleichung (2.12)

$f_0$	$\bar{\sigma}$ RMSDCF	$\sigma$	relerr%	min	max	ndata
0.05	0.00277	0.00011	4.05	0.00271	0.00297	5
0.1	0.00270	0.00014	5.12	0.00254	0.00285	5
0.15	0.00266	0.00009	3.39	0.00255	0.00280	5
0.2	0.00265	0.00007	2.67	0.00260	0.00277	5
0.25	0.00267	0.00006	2.30	0.00259	0.00276	5
0.3	0.00261	0.00012	4.54	0.00242	0.00271	5
0.4	0.00270	0.00012	4.42	0.00251	0.00282	5
0.5	0.00279	0.00007	2.41	0.00269	0.00287	5
0.6	0.00263	0.00014	5.36	0.00244	0.00277	5
0.7	0.00271	0.00010	3.67	0.00259	0.00287	5
0.8	0.00275	0.00013	4.85	0.00259	0.00294	5
0.9	0.00277	0.00011	4.05	0.00271	0.00297	5
$f_0$	$\bar{\sigma}$ RMSG	$\sigma$	relerr%	min	max	ndata
0.05	0.46880	0.00411	0.88	0.46449	0.47530	5
0.1	0.49062	0.00240	0.49	0.48650	0.49258	5
0.15	0.49693	0.00241	0.49	0.49288	0.49938	5
0.2	0.50128	0.00164	0.33	0.49909	0.50311	5
0.25	0.51087	0.00245	0.48	0.50838	0.51395	5
0.3	0.51379	0.00215	0.42	0.51095	0.51674	5
0.4	0.52179	0.00258	0.49	0.51857	0.52459	5
0.5	0.52708	0.00172	0.33	0.52493	0.52957	5
0.6	0.53165	0.00284	0.53	0.52875	0.53540	5
0.7	0.53478	0.00421	0.79	0.52863	0.53998	5
0.8	0.53879	0.00158	0.29	0.53759	0.54156	5
0.9	0.46880	0.00411	0.88	0.46449	0.47530	5

$f_0$	$\emptyset$ RMSDSN	$\sigma$	relerr%	min	max	ndata
0.05	0.18862	0.00045	0.24	0.18812	0.18925	5
0.1	0.18481	0.00058	0.32	0.18407	0.18562	5
0.15	0.18340	0.00023	0.13	0.18316	0.18371	5
0.2	0.18263	0.00022	0.12	0.18236	0.18296	5
0.25	0.18132	0.00033	0.18	0.18095	0.18179	5
0.3	0.18045	0.00031	0.17	0.18008	0.18078	5
0.4	0.17963	0.00043	0.24	0.17901	0.18019	5
0.5	0.17912	0.00037	0.20	0.17851	0.17941	5
0.6	0.17878	0.00013	0.07	0.17858	0.17893	5
0.7	0.17850	0.00016	0.09	0.17834	0.17876	5
0.8	0.17806	0.00025	0.14	0.17775	0.17837	5
0.9	0.18862	0.00045	0.24	0.18812	0.18925	5

Die grafische Auftragung in Abbildung 116 ergibt wiederum die gleichen Trends wie beim Heptan, niedrige  $f_0$ -Werte ergeben niedrige RMSG- sowie hohe RMSDSN-Werte. Die Überprüfung der optimierten Netzkarten auf Übersichtlichkeit ergibt keine schlüssigen Ergebnisse, da die Netzoberfläche bereits zu stark fragmentiert ist, um von großem Nutzen zu sein. Auf eine Abbildung wird verzichtet.



Abbildung 116: Variation von  $f_0$ , a) RMSDCF, b) RMSG, c) RMSDSN

### 2.8.5.1.3 Variation von $\Delta_{\max}(1)$

Tabelle 53: Variation von  $\Delta_{\max}(1)$  nach Gleichung (2.15)

$\Delta_{\max}(1)$	$\bar{\sigma}$ RMSDCF	$\sigma$	relerr%	min	max	ndata
10	0.00269	0.00008	2.96	0.00255	0.00273	5
9	0.00262	0.00009	3.41	0.00247	0.00269	5
8	0.00275	0.00012	4.42	0.00257	0.00287	5
7	0.00269	0.00015	5.41	0.00247	0.00286	5
6	0.00269	0.00009	3.32	0.00257	0.00280	5
5	0.00275	0.00006	2.32	0.00269	0.00285	5
4	0.00267	0.00012	4.48	0.00252	0.00283	5
3	0.00270	0.00009	3.35	0.00258	0.00282	5
2	0.00270	0.00013	4.85	0.00259	0.00289	5
1	0.00269	0.00008	2.88	0.00261	0.00282	5
$\Delta_{\max}(1)$	$\bar{\sigma}$ RMSG	$\sigma$	relerr%	min	max	ndata
10	0.49826	0.00236	0.47	0.49492	0.50125	5
9	0.50041	0.00169	0.34	0.49848	0.50266	5
8	0.50329	0.00146	0.29	0.50160	0.50547	5
7	0.50565	0.00167	0.33	0.50396	0.50797	5
6	0.50890	0.00217	0.43	0.50659	0.51144	5
5	0.51298	0.00238	0.46	0.50923	0.51535	5
4	0.51764	0.00297	0.57	0.51496	0.52202	5
3	0.52292	0.00269	0.52	0.51899	0.52530	5
2	0.53530	0.00063	0.12	0.53458	0.53608	5
1	0.65666	0.01365	2.08	0.64321	0.67879	5

$\Delta_{\max}(1)$	$\bar{\sigma}$ RMSDSN	$\sigma$	relerr%	min	max	ndata
10	0.18336	0.00025	0.14	0.18306	0.18371	5
9	0.18304	0.00023	0.12	0.18281	0.18334	5
8	0.18257	0.00034	0.19	0.18211	0.18306	5
7	0.18201	0.00022	0.12	0.18167	0.18223	5
6	0.18122	0.00034	0.18	0.18076	0.18159	5
5	0.18078	0.00015	0.09	0.18070	0.18106	5
4	0.17995	0.00011	0.06	0.17984	0.18007	5
3	0.17923	0.00019	0.10	0.17907	0.17954	5
2	0.17791	0.00023	0.13	0.17759	0.17820	5
1	0.17400	0.00036	0.20	0.17343	0.17441	5

Der gleiche Trend wie beim Hexan und Heptan bestätigt sich auch beim Oktan. Kleinere  $\Delta_{\max}(1)$ -Werte erzeugen große RMSG-Werte und umgekehrt. Die RMSDSN-Werte verhalten sich exakt gegenläufig.

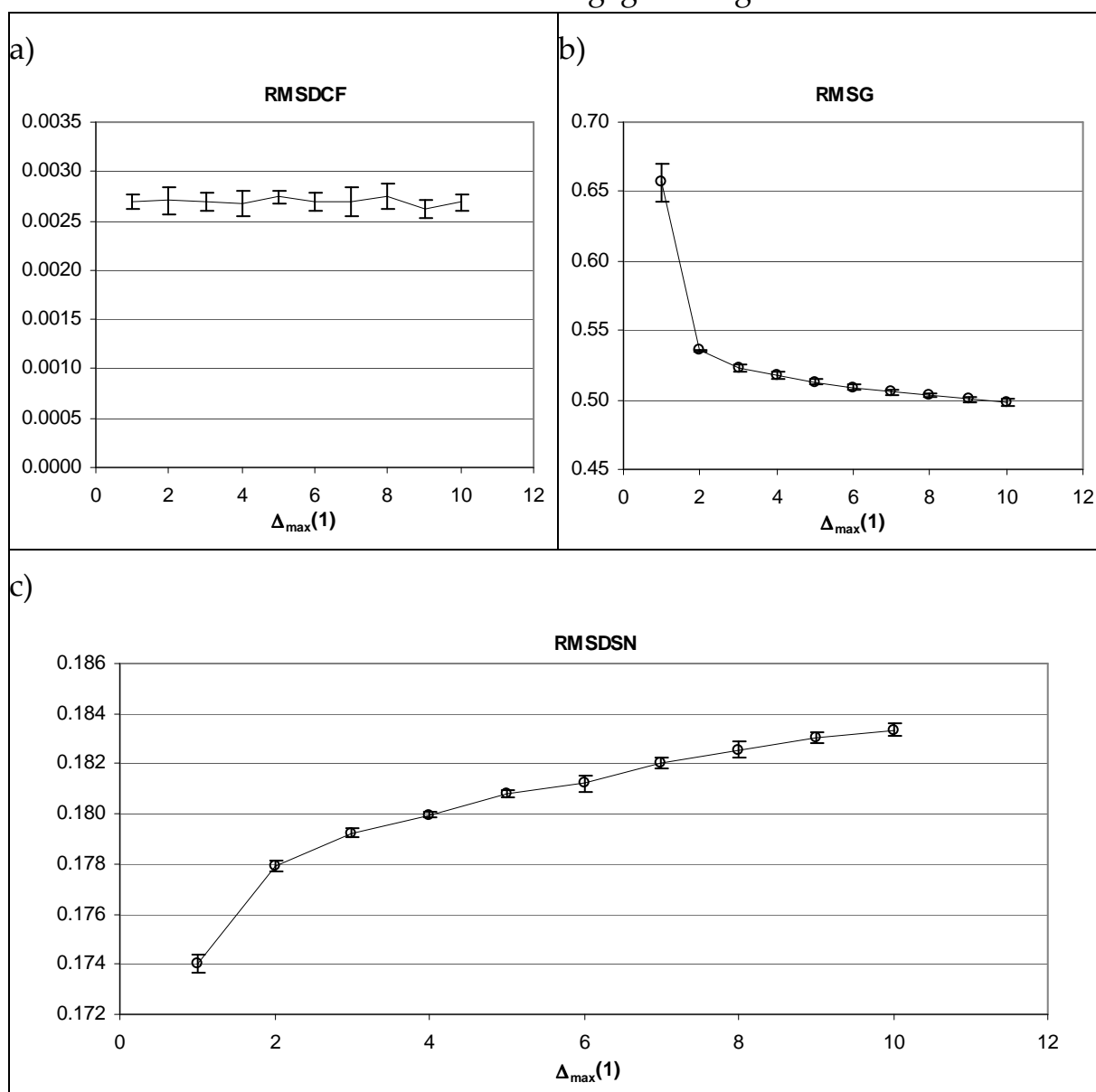


Abbildung 117: Variation von  $\Delta_{\max}(1)$ , a) RMSDCF, b) RMSG, c) RMSDSN

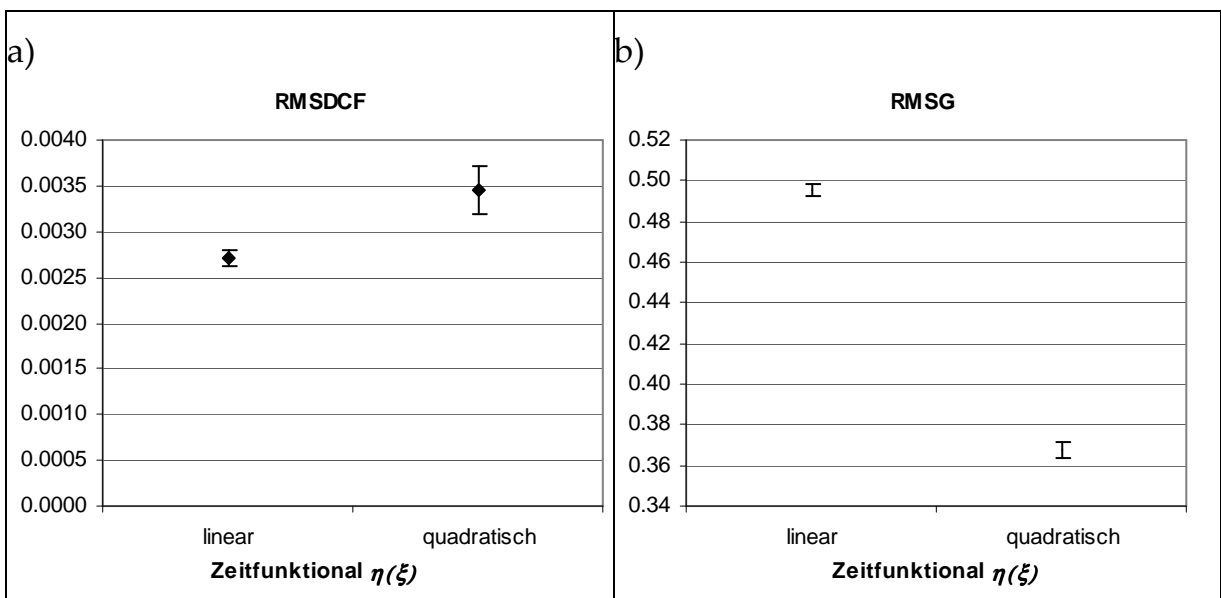


### 2.8.5.1.4 Variation von $\eta(\xi)$

Tabelle 54: Variation der Zeitabhängigkeit des Adaptionfaktors,  $\eta(\xi)$  nach Gleichung (2.17)

$\eta(\xi)$ (2.17)	$\emptyset$ RMSDCF	$\sigma$	relerr%	min	max	ndata
a) linear	0.00271	0.00009	3.2	0.00258	0.00280	5
b) quadratisch	0.00346	0.00026	7.5	0.00323	0.00385	5
$\eta(\xi)$ (2.17)	$\emptyset$ RMSG	$\sigma$	relerr%	min	max	ndata
a) linear	0.49549	0.003190	0.6	0.49083	0.49865	5
b) quadratisch	0.36769	0.003700	1.0	0.36218	0.37192	5
$\eta(\xi)$ (2.17)	$\emptyset$ RMSDSN	$\sigma$	relerr%	min	max	ndata
a) linear	0.18348	0.000443	0.2	0.18301	0.18402	5
b) quadratisch	0.20945	0.000403	0.2	0.20886	0.20996	5

Der Trend in Abbildung 118 ist wiederum identisch zum Heptan. Die RMSG-Werte sind kleiner für die quadratische Zeitfunktion, RMSDCF und RMSDSN kleiner für die lineare.



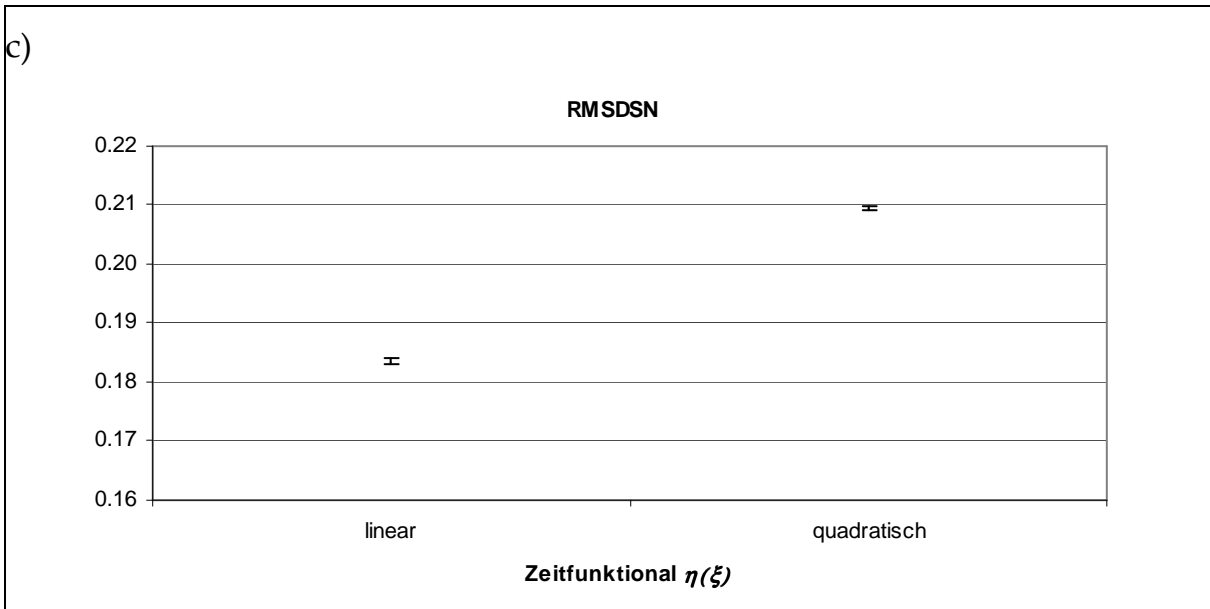


Abbildung 118: Variation von  $\eta(\xi)$  nach Gleichung (2.17)

### 2.8.5.1.5 Variation von NETX, NETY

Tabelle 55: Variation der Netzgröße NETX=NETY

NETX=NETY	$\bar{\sigma}$ RMSDCF	$\sigma$	relerr%	min	max	ndata
31	0.00169	0.00004	2.51	0.00165	0.00176	5
21	0.00272	0.00018	6.53	0.00252	0.00286	5
11	0.00541	0.00019	3.44	0.00517	0.00560	5
5	0.01229	0.00033	2.71	0.01202	0.01276	5
NETX=NETY	$\bar{\sigma}$ RMSG	$\sigma$	relerr%	min	max	ndata
31	0.40333	0.00201	0.50	0.40099	0.40570	5
21	0.49783	0.00214	0.43	0.49543	0.50091	5
11	0.67711	0.00334	0.49	0.67310	0.68000	5
5	0.99186	0.00365	0.37	0.98834	0.99737	5
NETX=NETY	$\bar{\sigma}$ RMSDSN	$\sigma$	relerr%	min	max	ndata
31	0.15539	0.00024	0.15	0.15521	0.15579	5
21	0.18338	0.00022	0.12	0.18307	0.18368	5
11	0.23993	0.00030	0.13	0.23968	0.24044	5
5	0.32810	0.00048	0.15	0.32749	0.32867	5

Die Auftragung der Werte in Abbildung 119 ergibt wiederum eine Verbesserung der Qualitätsparameter der Projektion für größere Werte.

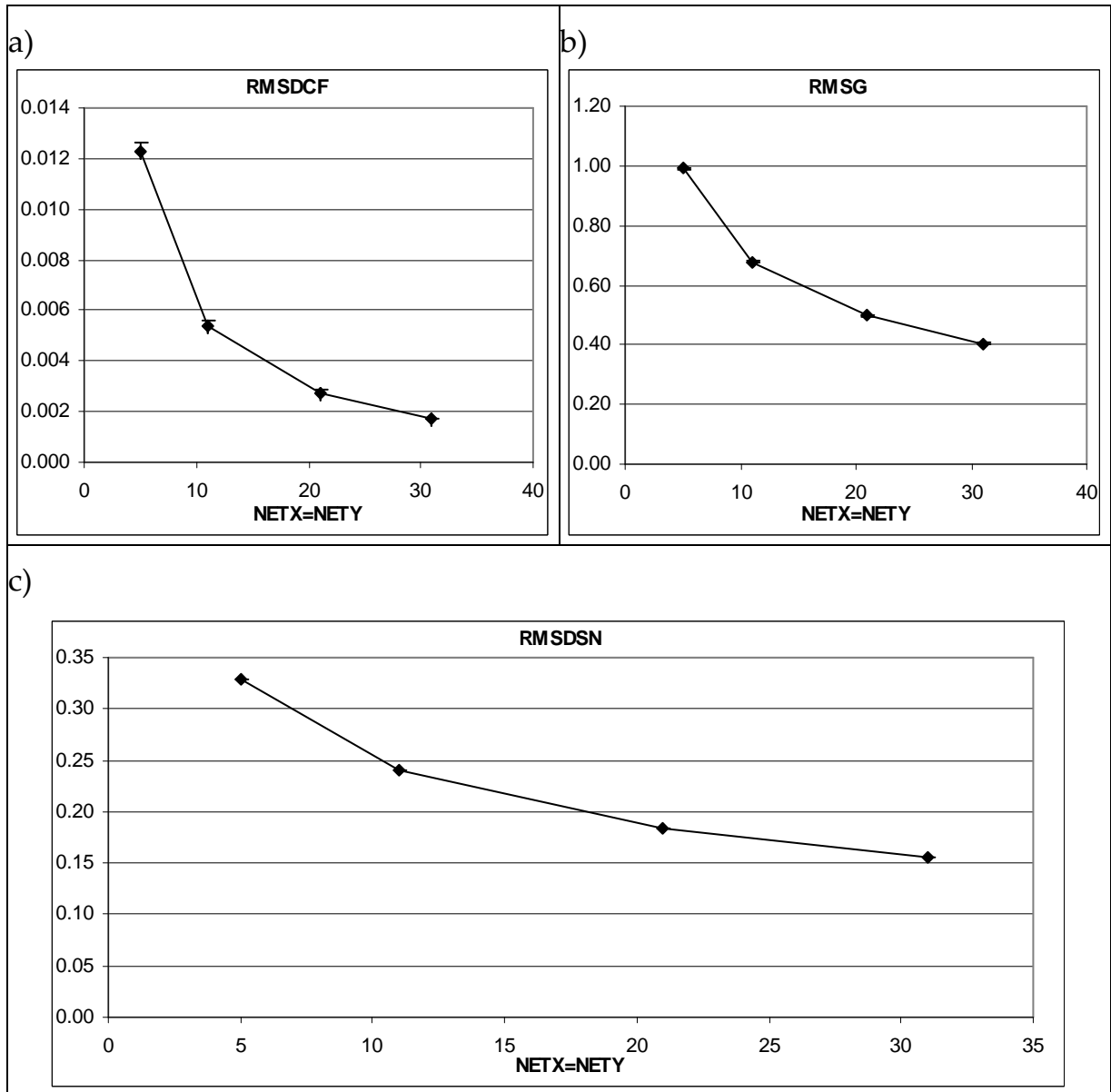


Abbildung 119: Variation der Netzgröße NETX=NETY

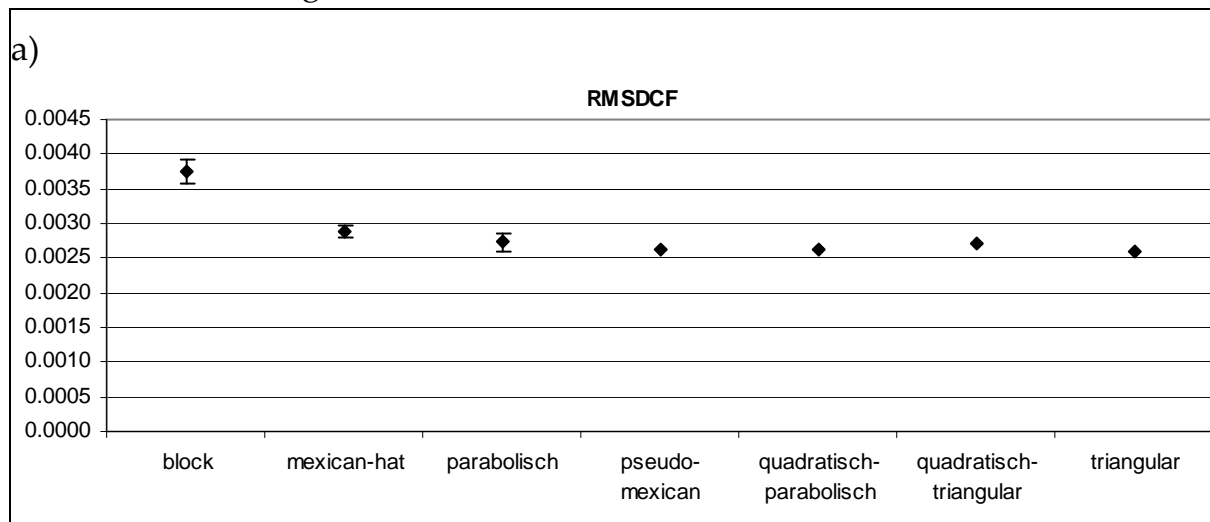
### 2.8.5.1.6 Variation der Abstandsfunktion $\vartheta(\zeta)$

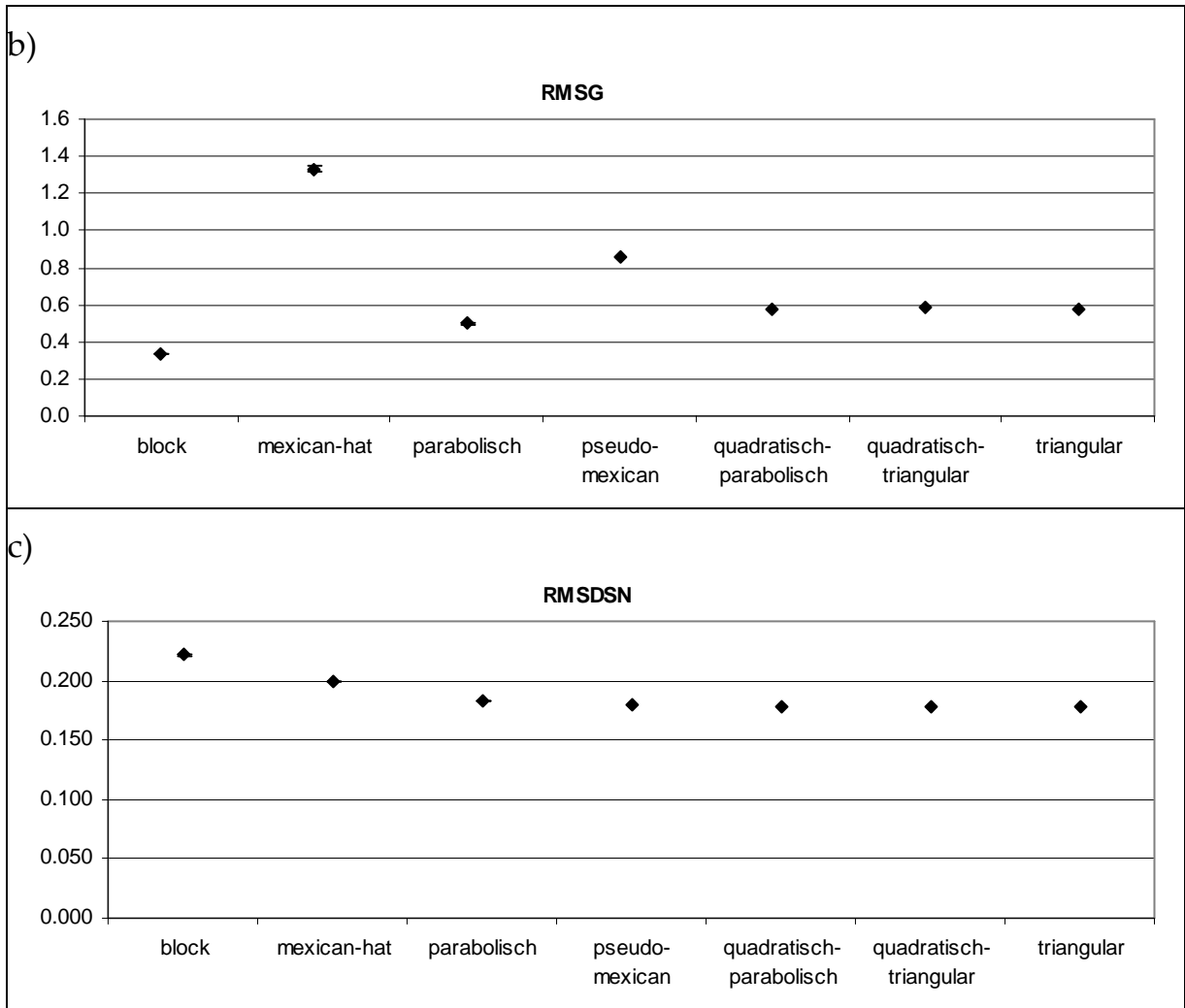
Tabelle 56: Variation der Abstandsfunktion  $\vartheta(\zeta)$  nach Gleichung (2.18)

$\vartheta(\zeta)$	$\bar{\vartheta}$ RMSDCF	$\sigma$	relerr%	min	max	ndata
<i>block</i>	0.00375	0.00017	4.51	0.00354	0.00394	5
<i>Mexican hat</i>	0.00288	0.00009	2.99	0.00273	0.00296	5
<i>parabolisch</i>	0.00273	0.00012	4.32	0.00258	0.00289	5
<i>pseudo Mexican</i>	0.00264	0.00022	8.45	0.00248	0.00302	5
<i>quadratisch-parabolisch</i>	0.00263	0.00007	2.69	0.00255	0.00274	5
<i>quadratisch-triangular</i>	0.00270	0.00010	3.53	0.00263	0.00287	5
<i>triangular</i>	0.00260	0.00013	5.19	0.00242	0.00272	5

$\vartheta(\zeta)$	$\emptyset$ RMSG	$\sigma$	relerr%	min	max	ndata
<i>block</i>	0.33611	0.00095	0.28	0.33499	0.33724	5
<i>Mexican hat</i>	1.33308	0.01737	1.30	1.31232	1.35906	5
parabolisch	0.49806	0.00223	0.45	0.49610	0.50133	5
<i>pseudo Mexican</i>	0.85496	0.01179	1.38	0.83621	0.86516	5
quadratisch- parabolisch	0.57822	0.00286	0.49	0.57373	0.58169	5
quadratisch- triangular	0.58050	0.00447	0.77	0.57666	0.58618	5
triangular	0.57838	0.00270	0.47	0.57513	0.58218	5
$\vartheta(\zeta)$	$\emptyset$ RMSDSN	$\sigma$	relerr%	min	max	ndata
<i>block</i>	0.22184	0.00080	0.36	0.22054	0.22245	5
<i>Mexican hat</i>	0.19912	0.00050	0.25	0.19831	0.19957	5
parabolisch	0.18323	0.00035	0.19	0.18267	0.18360	5
<i>pseudo Mexican</i>	0.17970	0.00040	0.22	0.17930	0.18035	5
quadratisch- parabolisch	0.17874	0.00020	0.11	0.17847	0.17903	5
quadratisch- triangular	0.17872	0.00032	0.18	0.17836	0.17918	5
triangular	0.17891	0.00021	0.12	0.17866	0.17923	5

Die aufgetragenen Qualitätsparameter ergeben den gleichen Trend wie beim Hexan und Heptan zuvor. Die RMSG- und die RMSDSN-/RMSDCF-Werte verhalten sich antagonistisch.





**Abbildung 120: Variation der Abstandsfunktion  $\vartheta(\zeta)$**

Die offensichtlichen Unterschiede im Verhalten der Abstandsfunktionen werfen wiederum die Frage auf, welche Auswirkungen die Qualitätsmerkmale auf die erhaltenen Netzarten im Einzelnen haben. Zu diesem Zweck werden in Abbildung 121 je eine Netzart der Einfärbung nach MINIX1 und MINIX2 unterworfen.

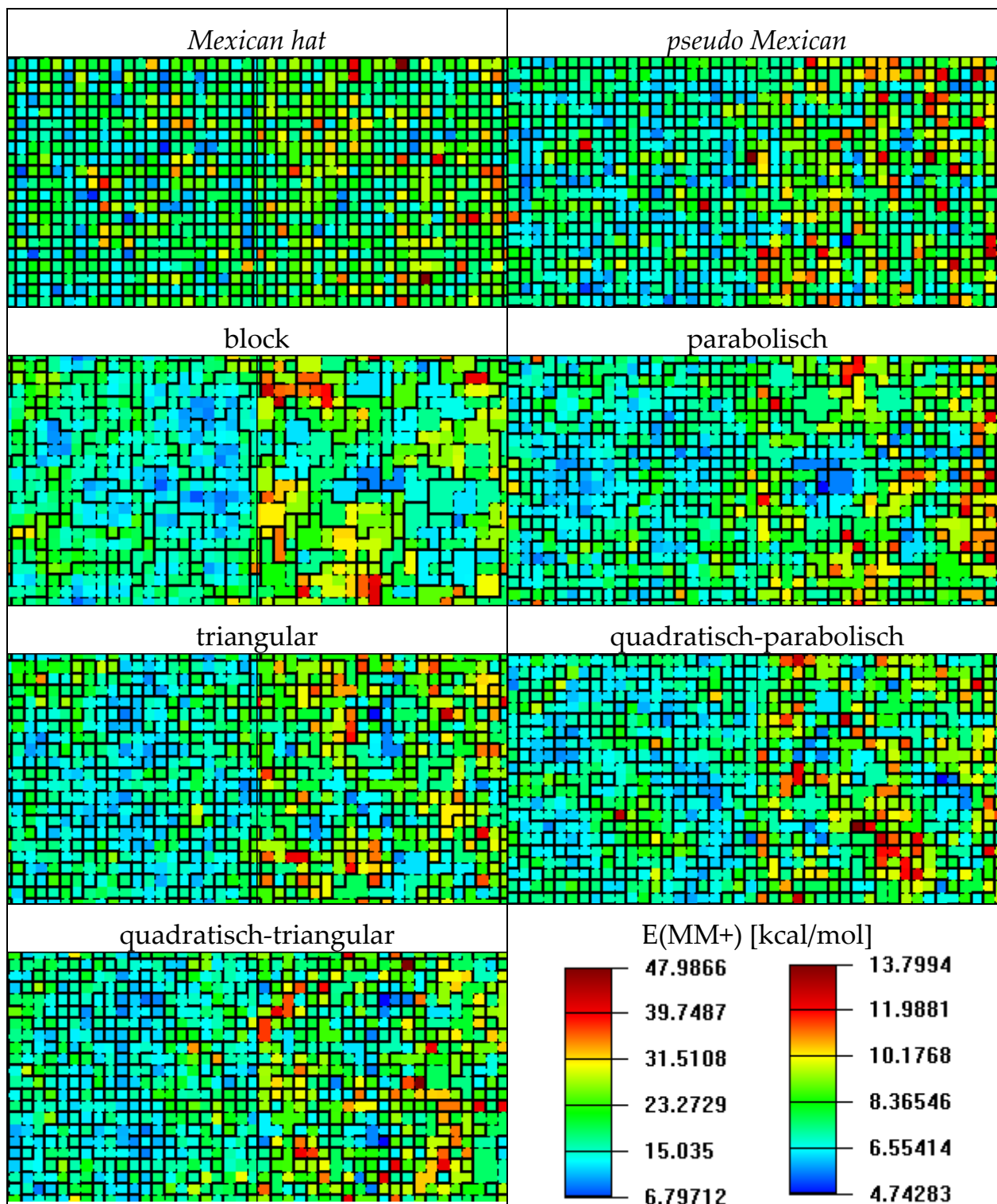


Abbildung 121: Beispielnetskarten Gridhyperfläche – Variation der Abstandsfunktion, Oktan, links partiell optimiert unter Verwendung der Netzrestraints nach COLORIX1, rechts vollständig optimiert nach COLORIX2

Das Ergebnis ist ausgeprägter als beim Heptan. Die „*Mexican hat*“-artigen Funktionen ergeben vollständig fragmentierte und wenig hilfreiche Netzkar-ten. Die übrigen Abstandsfunktionen fragmentieren in der Reihenfolge:

„*block*“ < „*parabolisch*“ < „*triangular*“ < „*quadratisch-parabolisch*“ < „*quadra-tisch-triangular*“. Der erhöhten Übersichtlichkeit der *block* Funktion steht eine

geringere Vollständigkeit der repräsentierten Konformationen gegenüber. So fehlt zum Beispiel ausgerechnet das globale *all-anti*-Minimum auf der *block*-Netzkarte und geht im hellblauen Bereich der *single-gauch*-Minima auf.

### 2.8.5.1.7 Fazit

Die Ergebnisse decken sich weitgehend mit den Ergebnissen, die beim Heptan erhalten wurden. Allerdings sind die erhaltenen Netzkarten zu unübersichtlich, um von großem Nutzen zum Verständnis des deterministisch ermittelten Konformationsraums des Oktanmoleküls zu sein. Es wurden folgende MTS-Werte (minimale Trainingsschritte nach Kapitel 2.5.4) ermittelt:

MTS(RMSG, 1%) > 7776000

MTS(RMSDSN, 1%) > 7776000

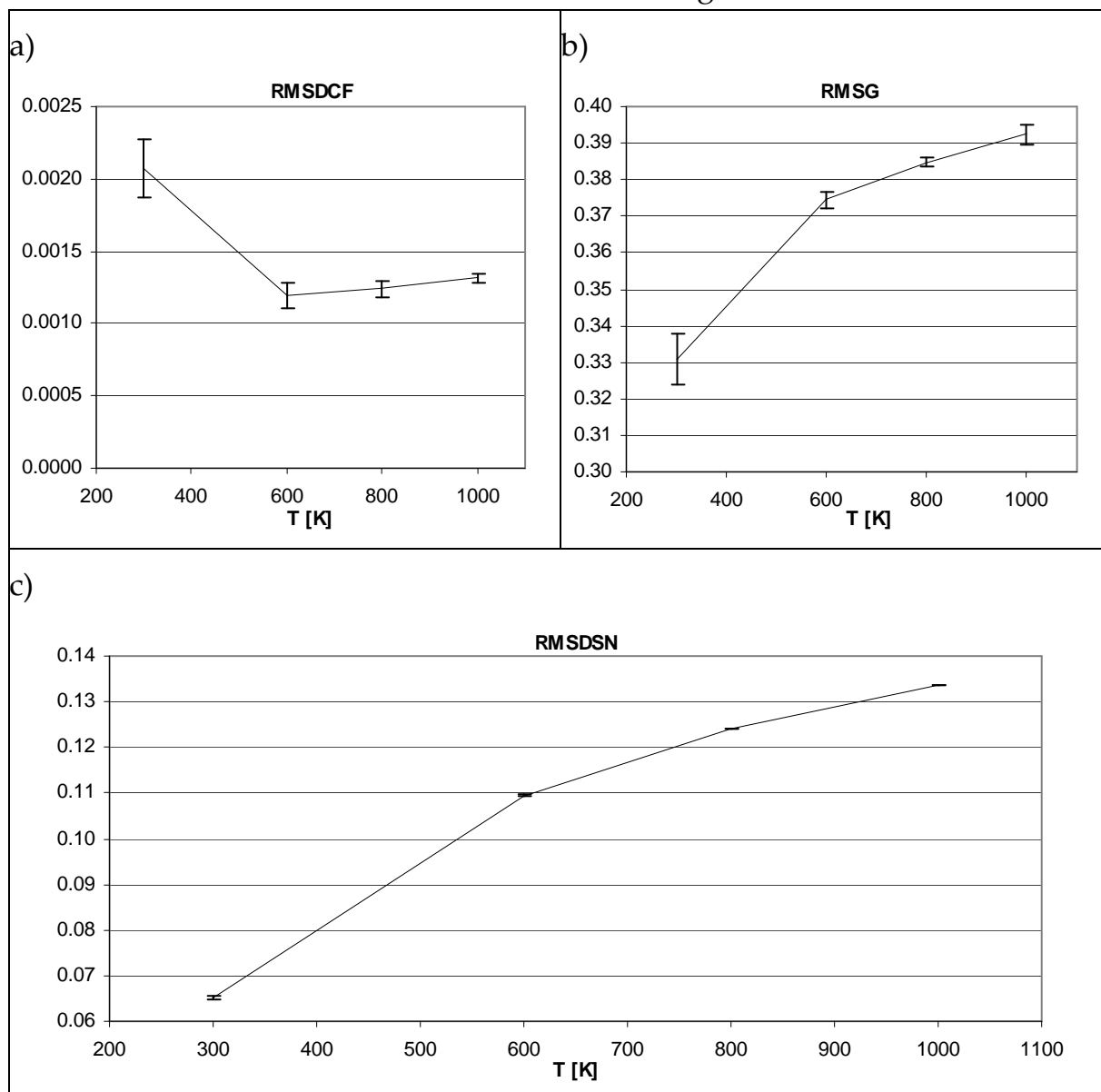
### 2.8.5.2 MD Oktan, Training der Diederwinkel

Für das Oktan-Molekül wird bei unterschiedlichen Temperaturen eine Gruppe Moleküldynamikrechnungen durchgeführt. (Kraftfeld HYPERCHEM MM+, in vacuo, 200000 Strukturen, Simulationsdauer 200000 ps, Aufruf von FIX\_ROTATION.EXE bei je 20 ps, *step size* 0.001 ps, *datafreq*=1/1000 *data steps*<sup>-1</sup>). Mit diesen MD-Eingabedaten werden neuronale Netze trainiert. [25x25 Neuronen; 100 *epochs*;  $\Delta_{\max}(1)=12$ ;  $f_0=0.4$ ; Abstandsfunktion parabolisch, Training der fünf CC-Diederwinkel]. Die erhaltenen Qualitätsparameter sind in Tabelle 57 angeführt. Eine Netzkarte (800 K) mit namentlich identifizierten und den Clustern der Oberfläche zugeordneten Konformationen ist bereits in Abbildung 114 angeführt.

**Tabelle 57: Oktan MD-Simulationen bei unterschiedlichen Temperaturen, Training der zentralen Diederwinkel**

T [K]	Ø RMSDCF	$\sigma$	relerr%	min	max	ndata
1000	0.00132	0.00003	2.46	0.00129	0.00137	5
800	0.00124	0.00005	4.17	0.00116	0.00129	5
600	0.00120	0.00009	7.34	0.00107	0.00129	5
300	0.00207	0.00020	9.66	0.00184	0.00229	5
T [K]	Ø RMSG	$\sigma$	relerr%	min	max	ndata
1000	0.39233	0.00256	0.65	0.38781	0.39394	5
800	0.38479	0.00111	0.29	0.38329	0.38627	5
600	0.37451	0.00220	0.59	0.37105	0.37681	5
300	0.33077	0.00700	2.11	0.32262	0.33862	5
T [K]	Ø RMSDSN	$\sigma$	relerr%	min	max	ndata
1000	0.13361	0.00010	0.08	0.13348	0.13373	5
800	0.12403	0.00013	0.10	0.12395	0.12425	5
600	0.10949	0.00019	0.17	0.10923	0.10969	5
300	0.06504	0.00033	0.51	0.06460	0.06540	5

Es zeigt sich der gleiche Trend für die RMSG- sowie RMSDSN-Werte wie beim Heptan. Mit steigender Temperatur steigt die Varianz der trainierten Eingabedaten und die RMSG- sowie RMSDSN-Werte steigen an.



**Abbildung 122: Variation der Temperatur – Oktan MD, Training der drei CC-Diederwinkel**

Exemplarische Netzarten nach Durchführung des MINIX1- und MINIX2-Moduls sind in Abbildung 123 wiedergegeben. Deutlich wird wiederum das Ansteigen des Anteils der Hochenergiekonformationen mit höheren Temperaturen. Außerdem ist erkennbar, dass die erhaltenen Netzarten deutlich weniger fragmentiert und damit übersichtlicher sind als vergleichbare Netzarten, die aus dem Training der synthetischen Hyperfläche erhalten werden, vergleiche Kapitel 2.8.5.1.6.



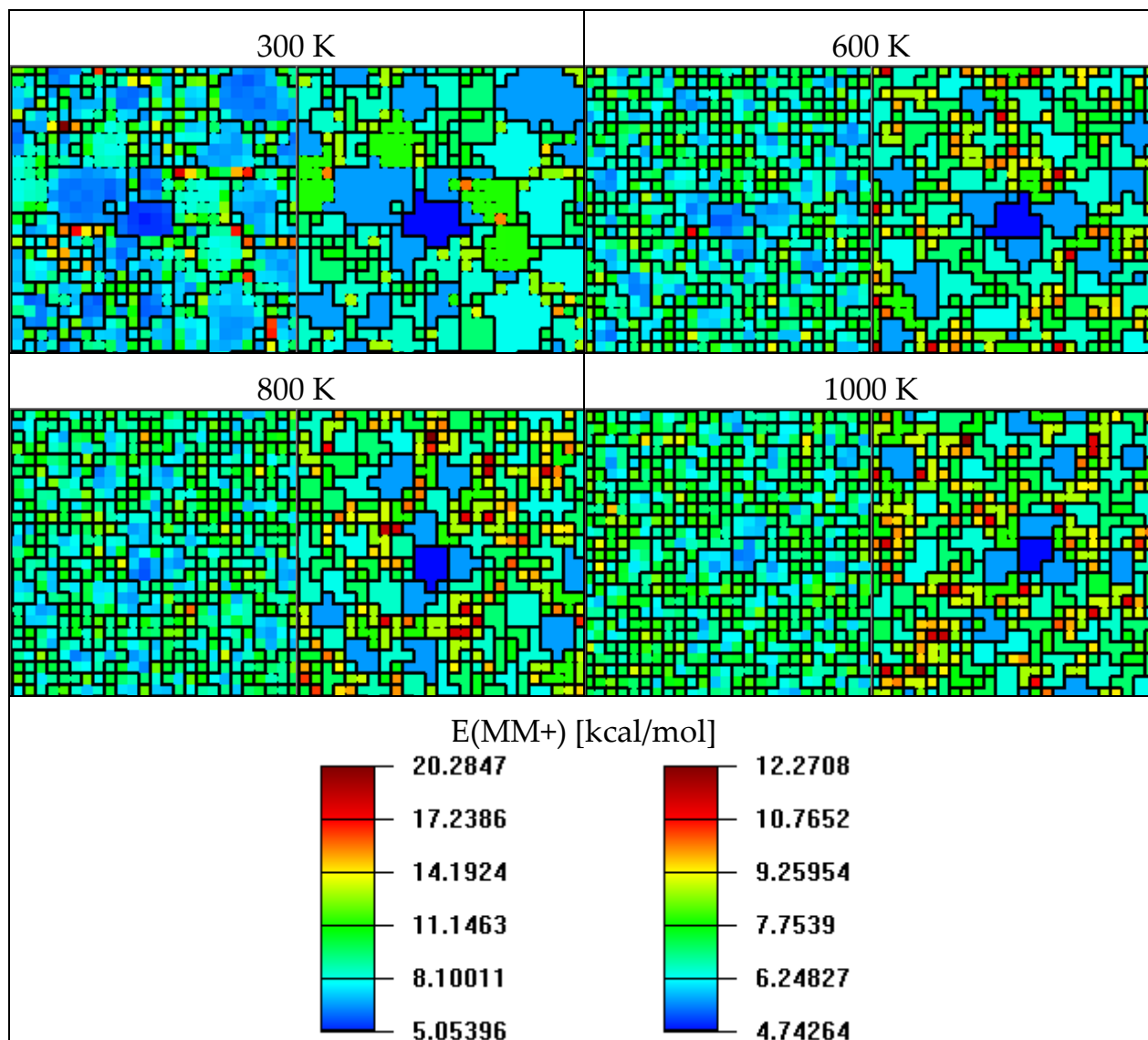


Abbildung 123: Beispielnetzkarten Gridhyperfläche – Variation der Temperatur – Oktan MD, Training der CC-Diederwinkel, links partiell optimiert unter Verwendung der Netzrestraints nach COLORIX1, rechts vollständig optimiert nach COLORIX2

### 2.8.5.3 MD Oktan, Training 51 zufälliger Distanzen

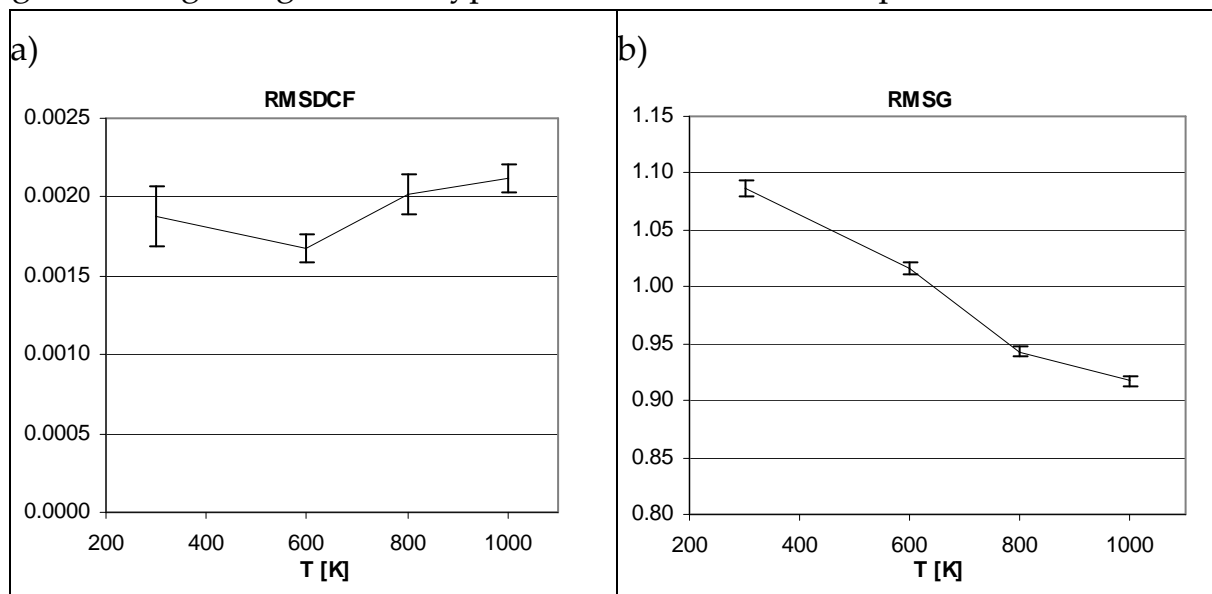
Aus Rechenzeitgründen wird repräsentativ für die komplette Abstandsmatrix ein Satz von einundfünfzig zufällig ausgewählten C-C, C-H und H-H Distanzen trainiert. Als Eingabedaten werden die gleichen MD-Rechnungen wie im vorherigen Kapitel verwendet.

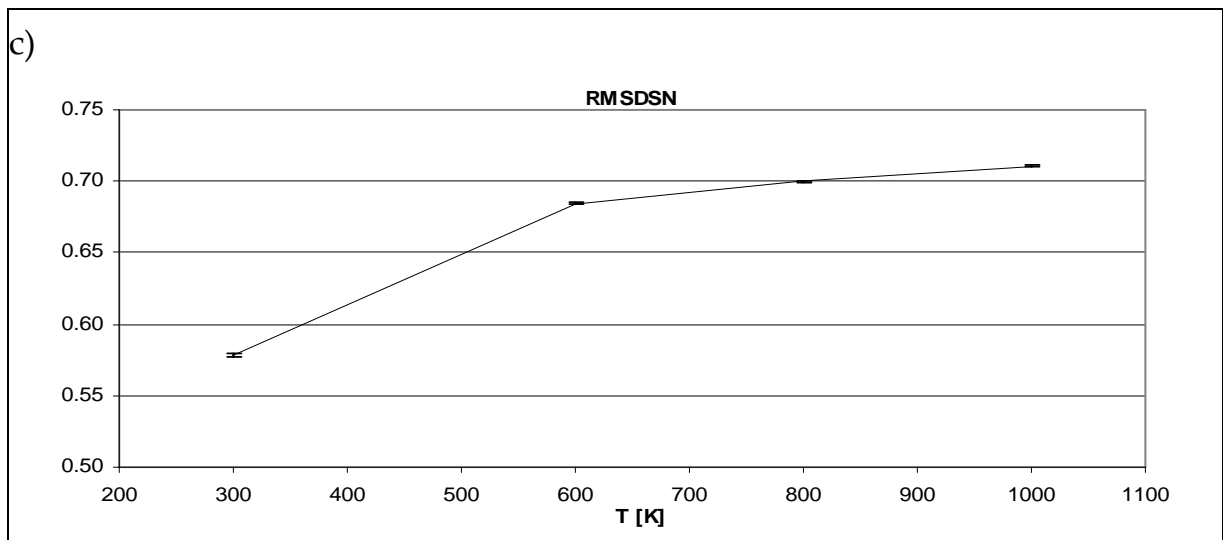
Die statistischen Daten des Netztrainings sind in Tabelle 58 aufgeführt. [25x25 Neuronen; 20 epochs;  $\Delta_{\max}(1)=12$ ;  $f_0=0.4$ ; Abstandsfunktion parabolisch, Training 51 zufällig ausgewählter Distanzen].

**Tabelle 58: Oktan MD-Simulationen bei unterschiedlichen Temperaturen, Training 42 zufällig ausgewählter Distanzen**

T [K]	Ø RMSDCF	$\sigma$	relerr%	min	max	ndata
1000	0.00212	0.00008	3.91	0.00198	0.00218	5
800	0.00202	0.00013	6.43	0.00185	0.00214	5
600	0.00168	0.00009	5.26	0.00157	0.00180	5
300	0.00188	0.00020	10.39	0.00159	0.00209	5
T [K]	Ø RMSG	$\sigma$	relerr%	min	max	ndata
1000	0.91698	0.00499	0.54	0.91050	0.92314	5
800	0.94307	0.00410	0.43	0.93631	0.94683	5
600	1.01617	0.00482	0.47	1.01103	1.02319	5
300	1.08714	0.00693	0.64	1.07931	1.09588	5
T [K]	Ø RMSDSN	$\sigma$	relerr%	min	max	ndata
1000	0.71073	0.00019	0.03	0.71051	0.71100	5
800	0.69969	0.00010	0.01	0.69956	0.69983	5
600	0.68427	0.00059	0.09	0.68337	0.68494	5
300	0.57864	0.00099	0.17	0.57758	0.58020	5

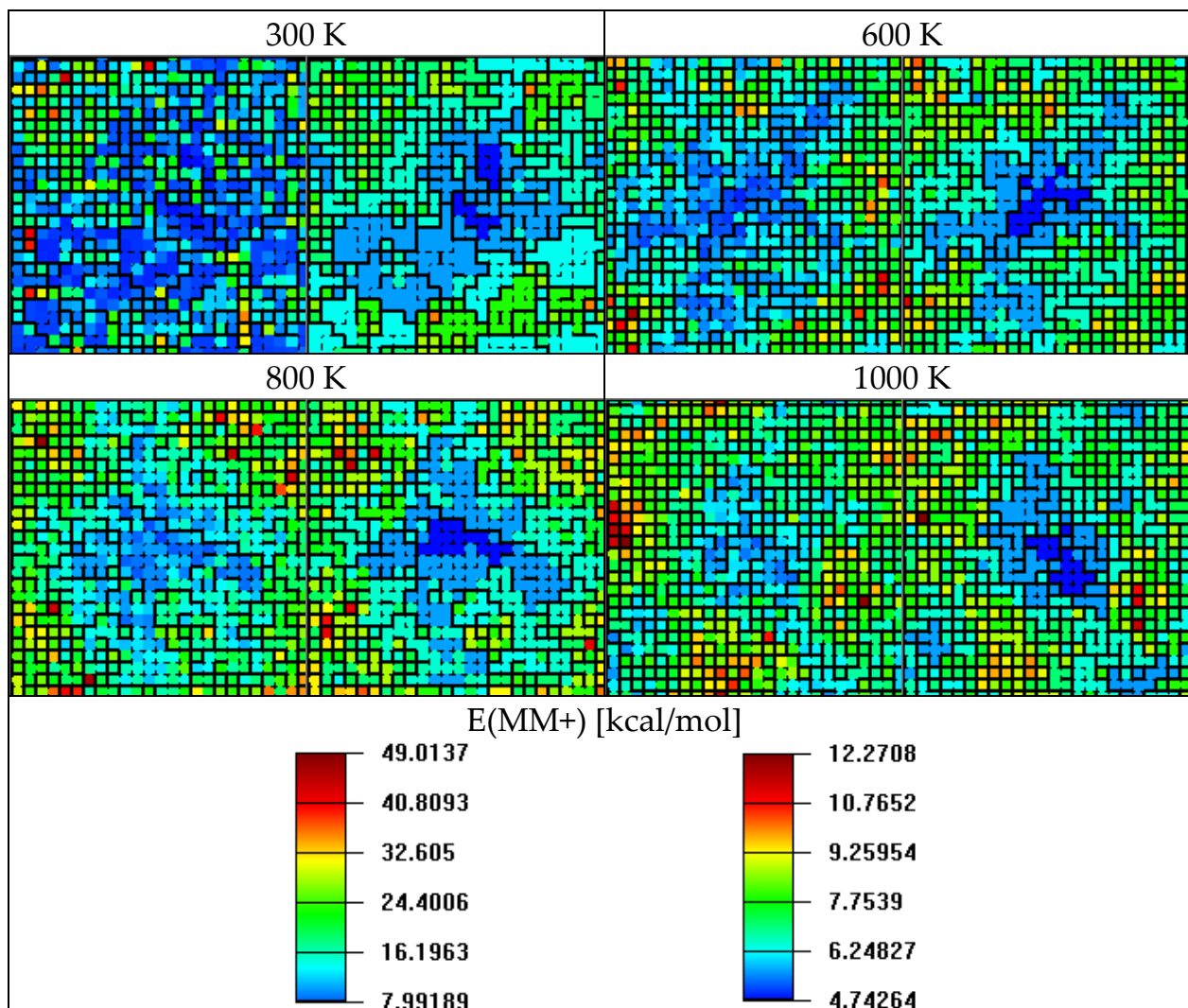
Im Gegensatz zum Training der Diederwinkel zeigen die Netzkarten, die durch Training zufälliger Distanzen erhalten werden, sinkende RMSG-Werte, siehe Abbildung 124. Dies lässt sich auf eine nicht wesentlich vergrößerte, aber gleichmäßiger abgedeckte Hyperfläche bei höheren Temperaturen erklären.





**Abbildung 124: Variation der Temperatur – Oktan MD, Training 51 zufällig ausgewählter Distanzen**

Obwohl die Zahl der trainierten internen Koordinaten und damit die Dimensionalität der Eingabedaten deutlich höher ist, werden vom Prinzip her einfache Netzoberflächen erhalten, siehe Abbildung 125.



**Abbildung 125:** Beispielnetzkarten Gridhyperfläche – Variation der Temperatur – Oktan MD, Training 42 zufällig ausgewählter Distanzen, links partiell optimiert unter Verwendung der Netzrestraints nach COLORIX1, rechts vollständig optimiert nach COLORIX2

#### 2.8.5.4 MD Oktan, Training der Schweratomdistanzen

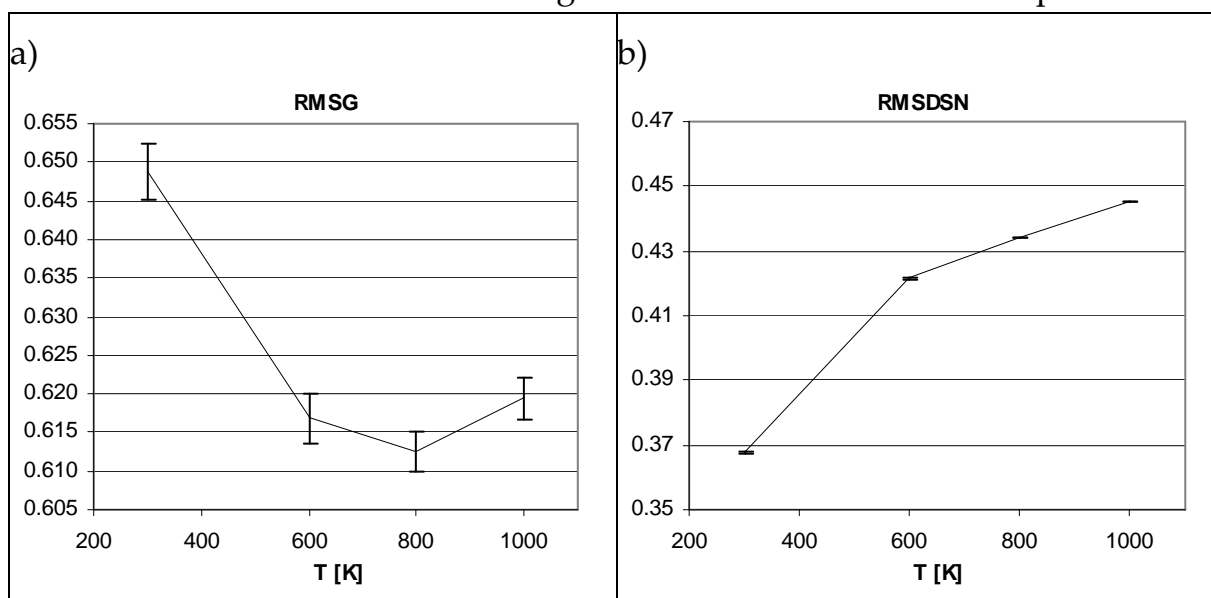
Das Training der 28 CC-Abstände führt wiederum zu einer vereinfachten Hyperfläche ohne Unterscheidung von Enantiomeren. Die statistischen Daten des Netztrainings [25x25 Neuronen; 20 *epochs*;  $\Delta_{\max}(1)=12$ ;  $f_0=0.4$ ; Abstandsfunktion parabolisch, Training der CC-Distanzen] sind in Tabelle 59 aufgeführt.

**Tabelle 59:** Oktan MD-Simulationen bei unterschiedlichen Temperaturen, Training der Schweratomdistanzmatrix

T [K]	Ø RMSG	$\sigma$	relerr%	min	max	ndata
1000	0.61944	0.00266	0.43	0.61547	0.62202	5
800	0.61254	0.00261	0.43	0.61001	0.61613	5
600	0.61683	0.00332	0.54	0.61240	0.62164	5
300	0.64874	0.00358	0.55	0.64487	0.65355	5

T [K]	Ø RMSDSN	$\sigma$	relerr%	min	max	ndata
1000	0.44537	0.00003	0.01	0.44534	0.44542	5
800	0.43400	0.00001	0.00	0.43399	0.43402	5
600	0.42148	0.00006	0.01	0.42138	0.42153	5
300	0.36778	0.00017	0.05	0.36757	0.36801	5

Die grafische Auftragung der obigen Tabellenwerte zeigt einen Abfall der RMSG-Werte bis 800 K mit folgendem Anstieg bei 1000 K. Die Cluster der Netzkarte für 800 K aus Abbildung 127 werden exemplarisch namentlich zugeordnet in Abbildung 114. Der trainierte Konformationsraum wird anscheinend bei höheren Temperaturen zunächst besser abgedeckt, was zu verringerten Neuron/Neuron-Gradienten führt. Ab einer gewissen Temperatur überwiegt dann der Effekt des insgesamt größer werdenden Konformationsraums. Die RMSDSN-Werte steigen kontinuierlich mit der Temperatur.



**Abbildung 126: Variation der Temperatur – Oktan MD, Training der Schweratomdistanzmatrix**

Durch das Wegfallen der Enantiomere erhält man eine relativ einfache Hyperfläche. Bei 300 K ist in Abbildung 127 außerdem zu beobachten, dass einige Konformationen, z.B. eine der hellblauen *single-gauche* Konformationen (aagaa und aaGaa), übermäßig stark populiert sind. Dies unterstützt die These, dass gleichmäßiger abgedeckte Hyperflächen (600 K - 1000 K) niedrigere RMSG-Werte zur Folge haben.

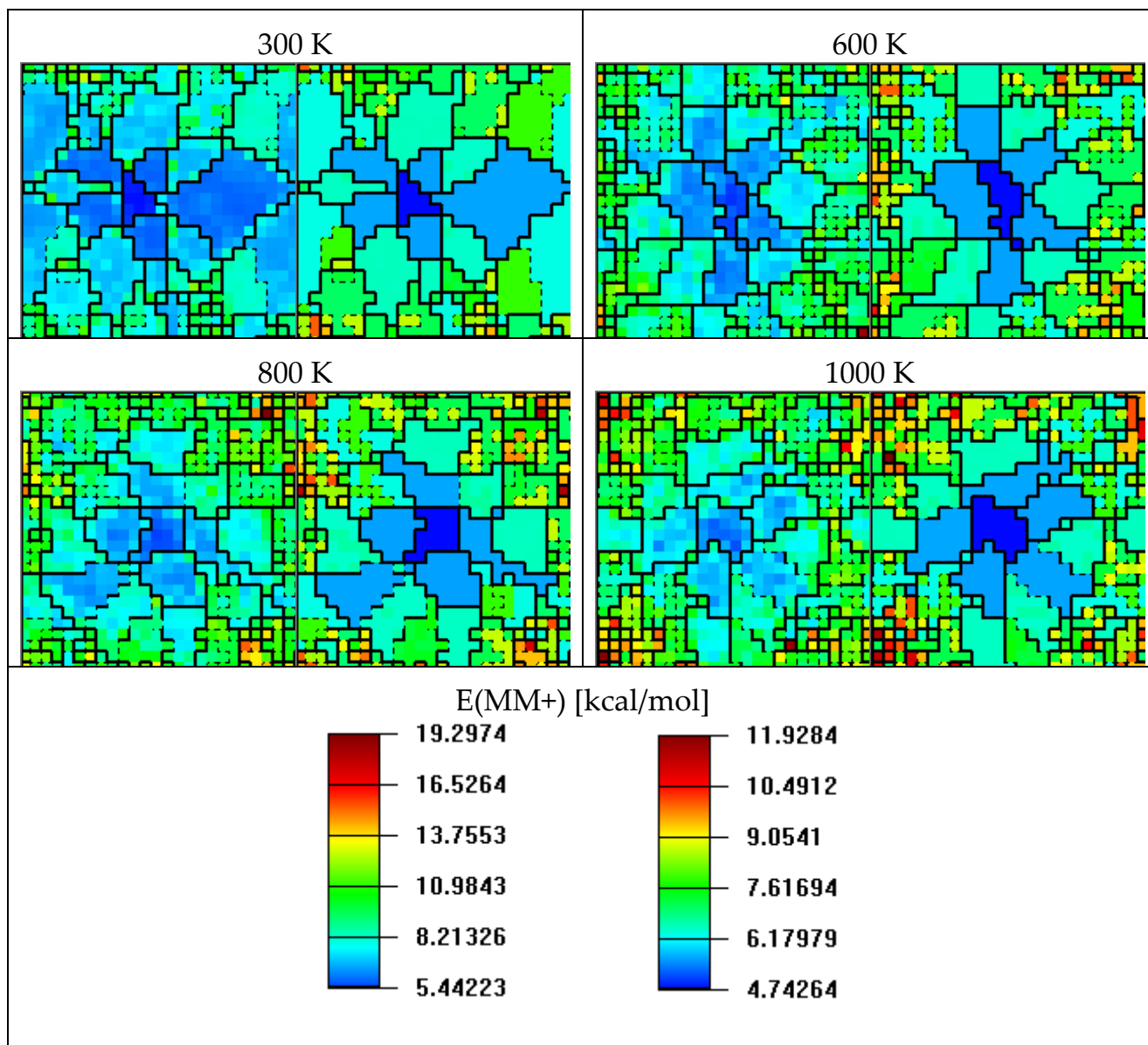


Abbildung 127: Beispielnetzarten Gridhyperfläche – Variation der Temperatur – Oktan MD, Training der Schweratomdistanzmatrix, links partiell optimiert unter Verwendung der Netzrestraints nach COLORIX1, rechts vollständig optimiert nach COLORIX2

### 2.8.6 Cyclohexan

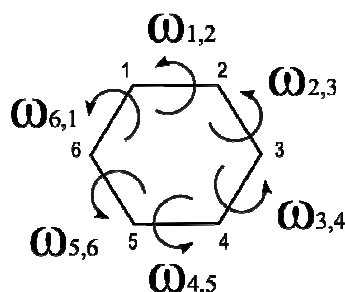


Abbildung 128: CC-Diederwinkel im Cyclohexan

Die Energiehyperfläche des Cyclohexan dient als Modell für die in zahlreichen Biomolekülen auftauchenden Sechsringe und ist experimentell gut untersucht. Eine Auftragung experimentell ermittelter Energien für Minima und Über-

gangszustände ist einem Lehrbuch für Stereochemie von J. Dale entnommen<sup>[111]</sup> und in Abbildung 129 wiedergegeben:

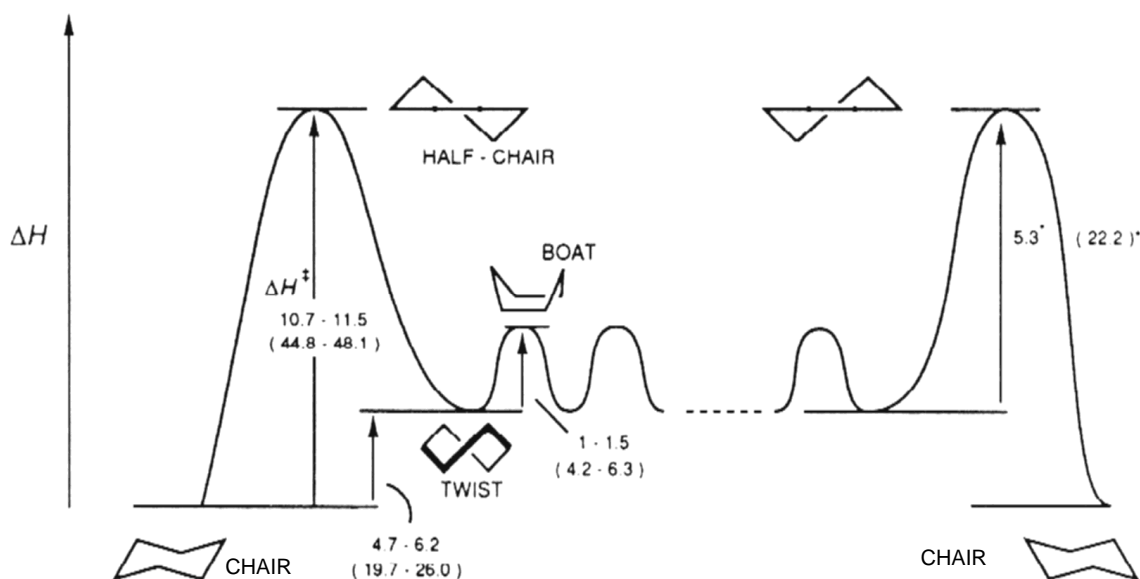


Abbildung 129: Energieprofil der Cyclohexan-Ringinversion nach J.Dale<sup>[111]</sup>

Eine neuere und topologisch richtigere mathematische Darstellung der Hyperfläche mittels zweier *ring-puckering* Koordinaten wurde von Hendrickson eingeführt<sup>[112]</sup> und von Cremer und Pople verfeinert.<sup>[113]</sup>

Folgerichtig lassen sich obige stationäre Punkte als Kugeloberfläche interpretieren, wie im Schema in Abbildung 130 angegeben. Die mit Zahlen bezeichneten Konformationen an den Polen sind dabei die *chair*-Minima und die mit Buchstaben bezeichneten Konformationen sind die *twist*-Minima. Die auf den „Polen“ liegenden *chair*-Minima sind mit jeder der *twist*-Minima über je einen spezifischen *half-chair*-Übergangszustand verbunden. Demgegenüber sind die auf dem „Äquator“ lokalisierten *twist*-Minima nur mit je zwei *twist*-Minima über je einen *boat*-Übergangszustand benachbart.

Auf der Kugel gegenüberliegende Konformationen sind dabei immer Enantiomere.

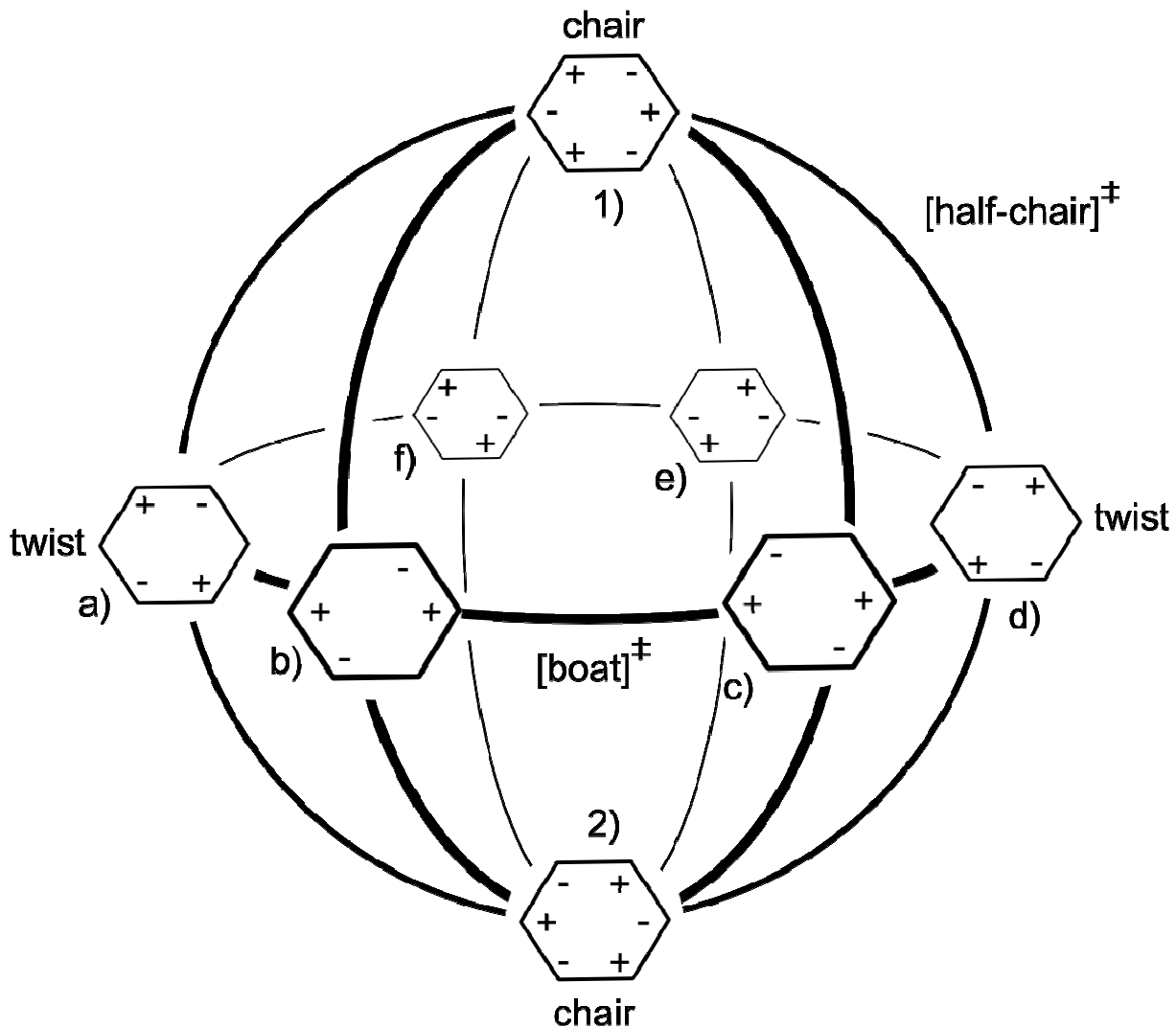


Abbildung 130: Konformationsübergänge im Cyclohexan nach Hendrickson, Pople und Cramer<sup>[112],[113]</sup>

C-Atome, die aus der Papierebene herausragen, sind mit einem Plus(+) bezeichnet, in die Papierebene eindringende C-Atome mit einem Minus(-). Die Verbindungslinien bezeichnen mögliche Konformationsübergänge (*half-chair* und *boat* Übergangszustände, siehe Abbildung 129). Übersetzt in die bereits verwendete *anti/gauche*-Schreibweise erhält man Tabelle 60. (Aufzählung der Atome in der Plus/Minus-Nomenklatur mit dem untersten Atom als Atom 1 und dann gegen den Uhrzeigersinn; der Unterstrich steht für ein Atom in der Papierebene).



**Tabelle 60: Übersetzung der in Abbildung 130 verwendeten Plus/Minus-Nomenklatur in die anti/gauche-Nomenklatur**

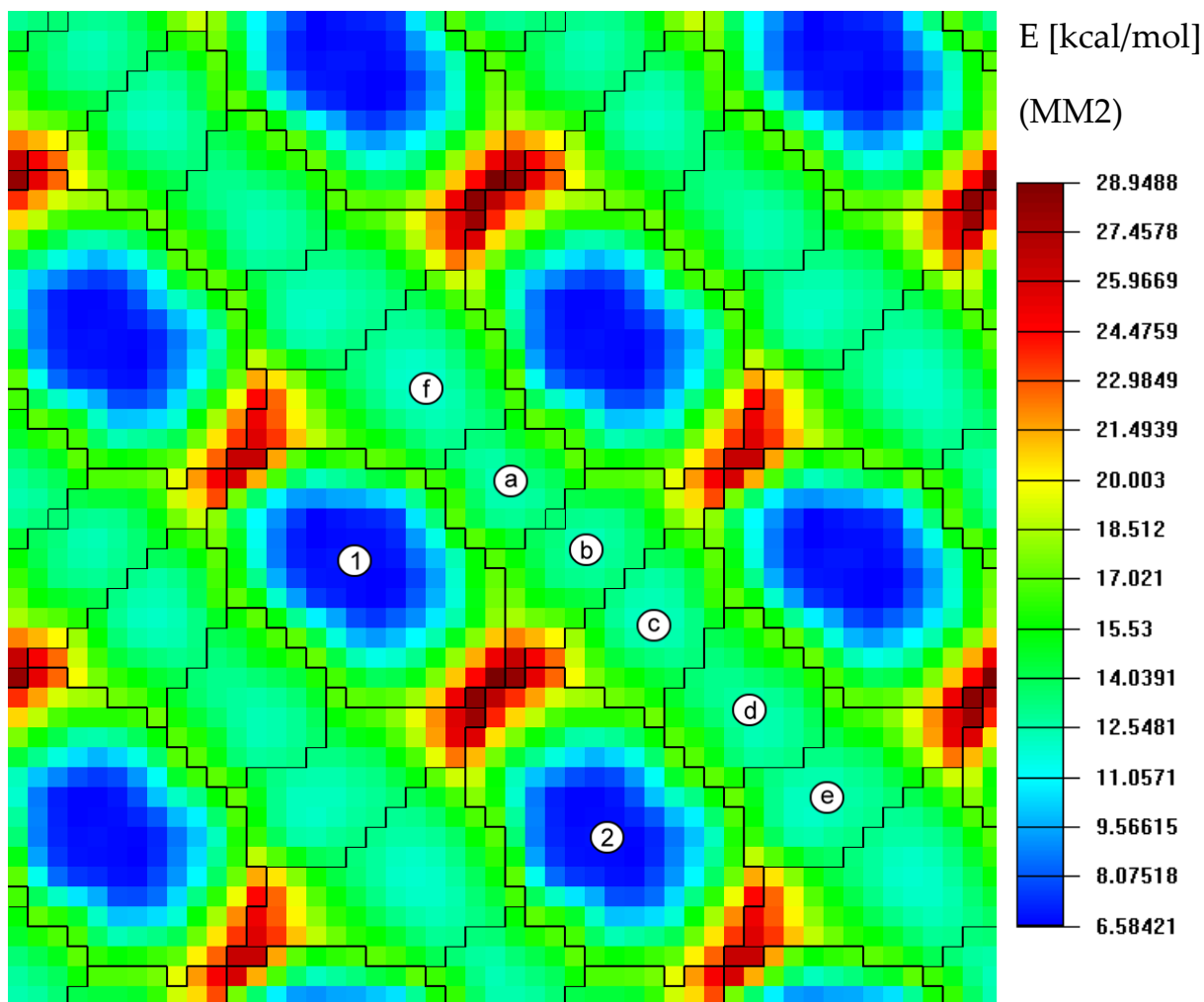
Konformation	Kurzname	Plus/Minus	Anti/gauche
<i>chair</i>	1	+--+ +-	gGgGgG
<i>chair</i>	2	-+ -+ -+	GgGgGg
<i>twist</i>	a	_+ _+ -	GggGgg
<i>twist</i>	b	+ _+ _-	GgGGgG
<i>twist</i>	c	+ -+ _-	ggGggG
<i>twist</i>	d	_+ _+ -	gGGgGG
<i>twist</i>	e	- _+ -+	gGggGg
<i>twist</i>	f	-+ _+ -	GGgGGg

### 2.8.6.1 Training der Grid-Hyperfläche

Der Konformationsraum des Cyclohexans wird für folgende Auswertungen per Gridrechnung mit Hilfe des Programms CYCHEX\_TORS.EXE, Kapitel 5.3.6 erzeugt (3366 Strukturen). In Kurzform: Die im Rahmen der Gridrechnung erwünschten Ringdiederwinkel werden als *restraints* gesetzt und das Cyclohexanmolekül einer Geometrieoptimierung unterzogen. Die Energien sind das Ergebnis einer nachfolgenden *restraint*-freien *single-point* Rechnung.

Das Training eines Kohonennetzes ergibt nach MINIX1/COLORIX1 eine Netzkarte, welche in 2x2 – Form aneinandergelagert dargestellt ist, so dass jedes Neuron einmal mit all seinen 8 direkten Nachbarn zu sehen ist (siehe Abbildung 131). [ $f_0=0.15$ ,  $epochs=1000$ ,  $neighborhood\ function =$  parabolisch,  $NETX=NETY=25$ ,  $linear\ learning\ decay$ ]

In den Strukturen der durch oben beschriebene Gridrechnung erhaltenen synthetischen Hyperfläche existiert auch ein vollständig planarer stationärer Punkt höherer Ordnung. Dieser würde sich in Abbildung 130 in der Mitte der Kugel befinden. Er entspricht dem direkten Übergang der beiden *chair*-Konformationen ineinander und ist in Abbildung 131 durch die dunkelroten Neuronen repräsentiert.



**Abbildung 131:** Kohonnennetz nach Training der Ringdiederwinkel der Konformationen aus oben beschriebener Gridrechnung des Cyclohexans. Um das Randproblem zu vermeiden sind vier Kohonnenkarten im Quadrat aneinandergesetzt dargestellt, so dass jede Konformation einmal inklusive der kompletten Nachbarschaft zu sehen ist.

Die Übergangszustände sind sämtlich topologisch richtig zwischen den Minima zu finden. Damit liefert die Kohonnenkarte eine anschaulichere grafische Repräsentation der Hyperfläche des Cyclohexans als bislang in der Literatur bekannt. Als dreidimensionale Auftragung mit Hilfe des Programms POV-Ray<sup>[114]</sup> in Abbildung 132 wird dies besonders deutlich. Dies ist umso erstaunlicher, weil die Hyperfläche der Cyclohexans wie bereits erwähnt von großem Interesse und bereits gut untersucht ist.

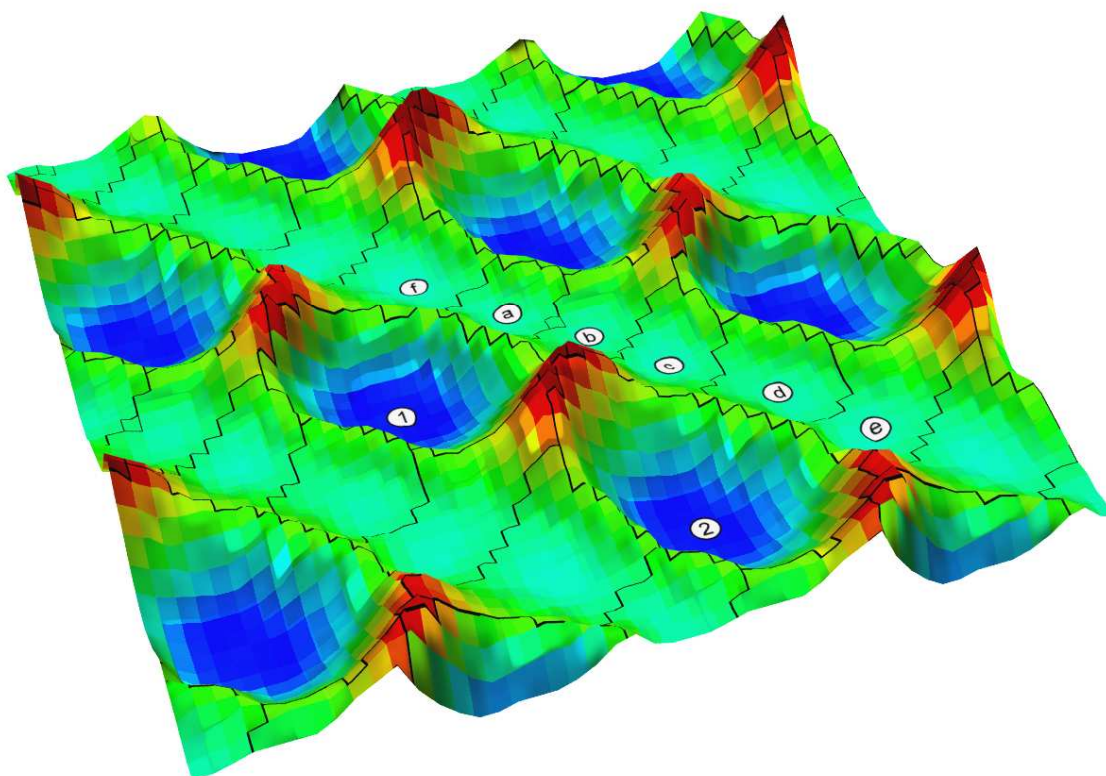


Abbildung 132: Dreidimensional gerenderte Kohonenkarte

Trainiert werden die Kohonnennetze in den folgenden Unterkapiteln mit Vorgabewerten von [ $f_0=0.15$ ,  $epochs=1000$ ,  $neighborhood\ function = \text{parabolisch}$ ,  $NETX=NETY=21$ ,  $linear\ learning\ decay$ ].

### 2.8.6.1.1 Variation der epochs

Tabelle 61: Cyclohexan - Variation der epochs

<i>epochs</i>	$\bar{\sigma}$ RMSDCF	$\sigma$	relerr%	min	max	ndata
1000	0.00164	0.00012	7.54	0.00147	0.00176	5
500	0.00190	0.00017	8.97	0.00170	0.00214	5
300	0.00182	0.00015	8.17	0.00169	0.00205	5
200	0.00194	0.00021	11.05	0.00165	0.00223	5
100	0.00195	0.00024	12.56	0.00160	0.00222	5
50	0.00214	0.00015	7.21	0.00194	0.00235	5
10	0.00198	0.00013	6.38	0.00185	0.00217	5
2	0.00207	0.00024	11.78	0.00179	0.00240	5
1	0.00215	0.00034	16.04	0.00184	0.00261	5

<i>epochs</i>	Ø RMSG	$\sigma$	relerr%	min	max	ndata
1000	0.14141	0.00077	0.54	0.14034	0.14205	5
500	0.13974	0.00063	0.45	0.13918	0.14072	5
300	0.13780	0.00024	0.18	0.13761	0.13816	5
200	0.13547	0.00042	0.31	0.13494	0.13591	5
100	0.13238	0.00062	0.47	0.13143	0.13300	5
50	0.12972	0.00035	0.27	0.12926	0.13016	5
10	0.12394	0.00026	0.21	0.12375	0.12440	5
2	0.11541	0.00088	0.76	0.11455	0.11680	5
1	0.11210	0.00117	1.05	0.11007	0.11291	5
<i>epochs</i>	Ø RMSDSN	$\sigma$	relerr%	min	max	ndata
1000	0.02485	0.00021	0.85	0.02459	0.02518	5
500	0.02536	0.00010	0.41	0.02522	0.02548	5
300	0.02595	0.00012	0.46	0.02577	0.02607	5
200	0.02687	0.00007	0.27	0.02675	0.02694	5
100	0.02891	0.00010	0.34	0.02876	0.02901	5
50	0.03094	0.00009	0.28	0.03088	0.03110	5
10	0.03697	0.00019	0.51	0.03680	0.03728	5
2	0.04748	0.00037	0.78	0.04701	0.04804	5
1	0.05323	0.00095	1.79	0.05217	0.05466	5

In Abbildung 133 wird die grafische Auftragung der obigen Tabellenwerte vorgenommen.  $\text{RMSG}_{\text{end}}^{[115]}$  und  $\text{RMSDSN}_{\text{end}}^{[116]}$  werden extrapoliert mit Programm Origin<sup>[105]</sup> für alle *epochs*>10. Bei keinem der Trainingsvorgänge werden genügend *epochs* trainiert, um das Netztraining zu „beenden“, sprich auf 1% an den finalen Wert heranzuführen. Der Verlauf der Kurven ist identisch zu den Molekülen Hexan...Oktan.

$$\text{RMSG}_{\text{end}} = 0.14579 \pm 0.00224, \quad \text{epoch}_{\text{end}}(\text{RMSG}) > 1000$$

$$\text{RMSDSN}_{\text{end}} = 0.02394 \pm 0.00046 \quad \text{epoch}_{\text{end}}(\text{RMSG}) > 1000$$

- $\text{MTS}(\text{RMSG}, 1\%) > 3366000$
- $\text{MTS}(\text{RMSDSN}, 1\%) > 3366000$

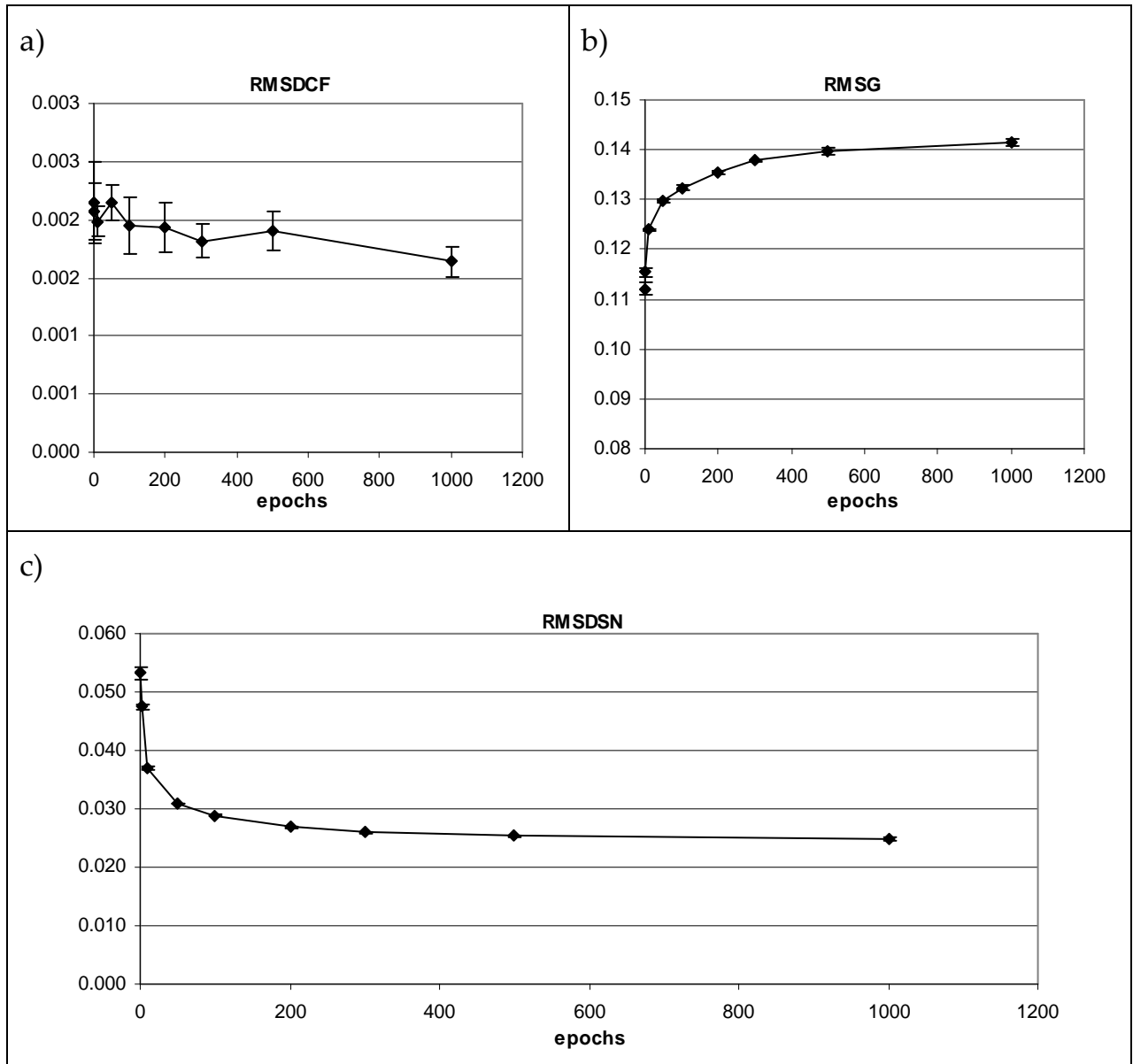


Abbildung 133: Variation der *epochs*, a) RMSDCF, b) RMSG, c) RMSDSN

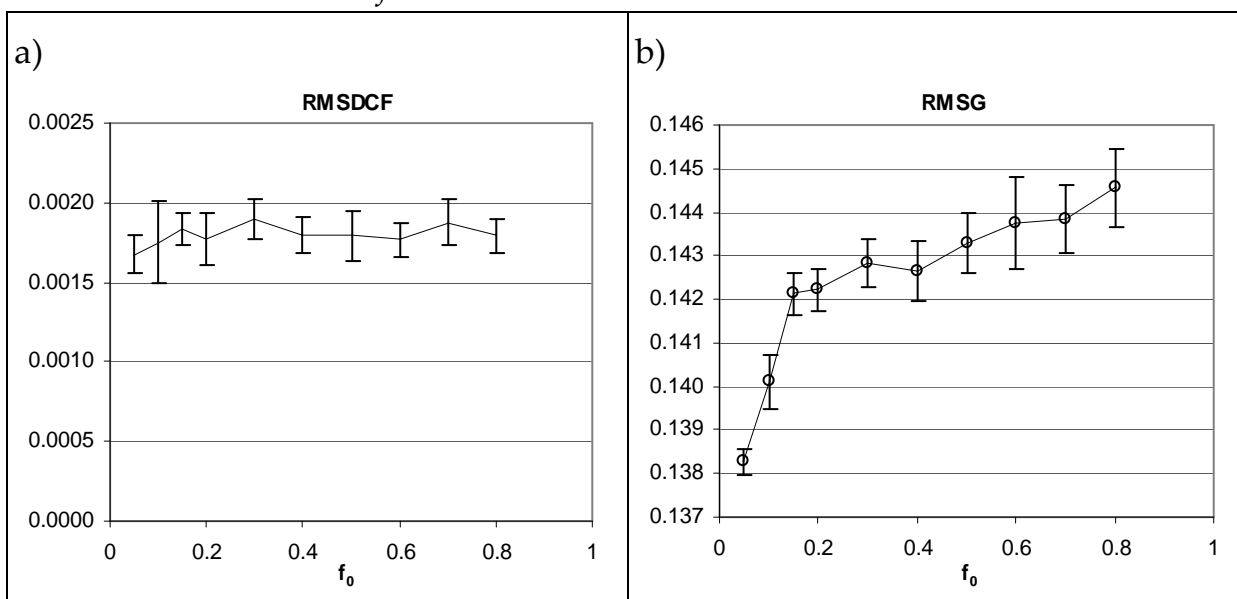
### 2.8.6.1.2 Variation von $f_0$

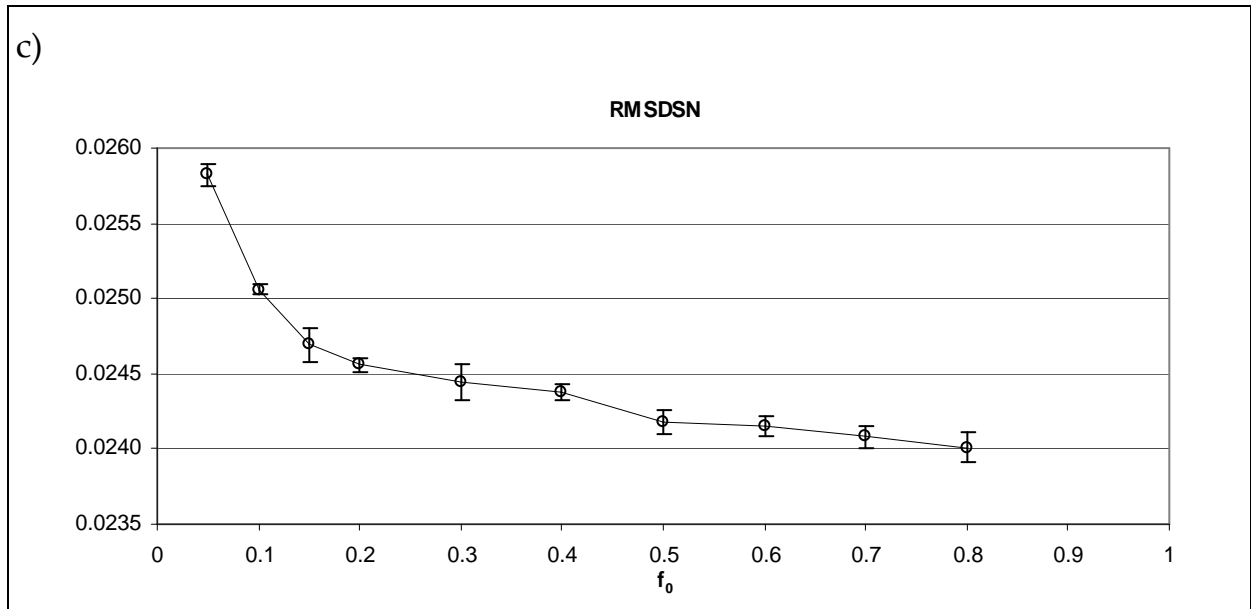
Tabelle 62: Variation des maximalen Adaptionfaktors  $f_0$  nach Gleichung (2.12)

$f_0$	$\bar{\sigma}$ RMSDCF	$\sigma$	relerr%	min	max	ndata
0.05	0.00168	0.00011	6.81	0.00154	0.00178	5
0.1	0.00175	0.00026	14.82	0.00151	0.00214	5
0.15	0.00184	0.00010	5.34	0.00175	0.00200	5
0.2	0.00177	0.00017	9.40	0.00159	0.00204	5
0.25	0.00189	0.00013	6.63	0.00175	0.00202	5
0.3	0.00180	0.00011	6.08	0.00167	0.00192	5
0.4	0.00179	0.00015	8.59	0.00161	0.00201	5
0.5	0.00177	0.00011	6.02	0.00163	0.00188	5
0.6	0.00187	0.00014	7.68	0.00172	0.00211	5
0.7	0.00179	0.00011	6.03	0.00162	0.00192	5
0.8	0.00168	0.00011	6.81	0.00154	0.00178	5
0.9	0.00175	0.00026	14.82	0.00151	0.00214	5

$f_0$	$\varnothing$ RMSG	$\sigma$	relerr%	min	max	ndata
0.05	0.13827	0.00029	0.21	0.13792	0.13859	5
0.1	0.14011	0.00062	0.44	0.13954	0.14114	5
0.15	0.14213	0.00049	0.34	0.14169	0.14292	5
0.2	0.14222	0.00048	0.34	0.14182	0.14299	5
0.25	0.14285	0.00055	0.38	0.14232	0.14362	5
0.3	0.14265	0.00068	0.48	0.14191	0.14364	5
0.4	0.14328	0.00070	0.49	0.14241	0.14391	5
0.5	0.14374	0.00105	0.73	0.14207	0.14490	5
0.6	0.14386	0.00078	0.54	0.14278	0.14461	5
0.7	0.14457	0.00089	0.62	0.14367	0.14603	5
0.8	0.13827	0.00029	0.21	0.13792	0.13859	5
0.9	0.14011	0.00062	0.44	0.13954	0.14114	5
$f_0$	$\varnothing$ RMSDSN	$\sigma$	relerr%	min	max	ndata
0.05	0.02582	0.00008	0.30	0.02573	0.02593	5
0.1	0.02506	0.00004	0.15	0.02501	0.02511	5
0.15	0.02469	0.00011	0.46	0.02453	0.02485	5
0.2	0.02456	0.00005	0.19	0.02451	0.02463	5
0.25	0.02444	0.00012	0.48	0.02430	0.02462	5
0.3	0.02437	0.00005	0.22	0.02428	0.02442	5
0.4	0.02418	0.00008	0.33	0.02409	0.02429	5
0.5	0.02415	0.00006	0.27	0.02408	0.02422	5
0.6	0.02408	0.00008	0.32	0.02395	0.02415	5
0.7	0.02401	0.00010	0.41	0.02392	0.02418	5
0.8	0.02582	0.00008	0.30	0.02573	0.02593	5
0.9	0.02506	0.00004	0.15	0.02501	0.02511	5

Die grafische Auftragung in Abbildung 134 ergibt wiederum die gleichen Trends wie beim Oktan mit niedrigen RMSG- und großen RMSDSN-Werten für kleine  $f_0$  Werte und umgekehrt. Auffällig ist der bereits zuvor beobachtete Knick in der Kurve bei  $f_0=0.15$ .



Abbildung 134: Variation von  $f_0$ , a) RMSDCF, b) RMSG, c) RMSDSN

### 2.8.6.1.3 Variation von $\Delta_{\max}(1)$

Tabelle 63: Variation von  $\Delta_{\max}(1)$  nach Gleichung (2.15)

$\Delta_{\max}(1)$	$\bar{\sigma}$ RMSDCF	$\sigma$	relerr%	min	max	ndata
10	0.00166	0.00014	8.70	0.00148	0.00179	5
9	0.00186	0.00015	7.94	0.00169	0.00203	5
8	0.00166	0.00016	9.71	0.00138	0.00179	5
7	0.00178	0.00012	6.95	0.00163	0.00193	5
6	0.00180	0.00018	10.18	0.00156	0.00203	5
5	0.00186	0.00016	8.47	0.00171	0.00209	5
4	0.00175	0.00020	11.49	0.00145	0.00196	5
3	0.00172	0.00015	8.79	0.00157	0.00190	5
2	0.00161	0.00011	6.83	0.00148	0.00172	5
1	0.00246	0.00084	34.05	0.00110	0.00324	5
$\Delta_{\max}(1)$	$\bar{\sigma}$ RMSG	$\sigma$	relerr%	min	max	ndata
10	0.14092	0.00052	0.37	0.14056	0.14178	5
9	0.14178	0.00049	0.35	0.14107	0.14237	5
8	0.14179	0.00048	0.34	0.14105	0.14221	5
7	0.14186	0.00036	0.25	0.14141	0.14236	5
6	0.14239	0.00070	0.49	0.14144	0.14342	5
5	0.14242	0.00049	0.35	0.14201	0.14324	5
4	0.14285	0.00031	0.22	0.14247	0.14327	5
3	0.14667	0.00131	0.89	0.14484	0.14807	5
2	0.14797	0.00221	1.49	0.14638	0.15185	5
1	0.21895	0.00717	3.27	0.21246	0.23049	5

$\Delta_{\max}(1)$	$\bar{\sigma}$ RMSDSN	$\sigma$	relerr%	min	max	ndata
10	0.02493	0.00016	0.65	0.02468	0.02511	5
9	0.02474	0.00012	0.47	0.02465	0.02493	5
8	0.02468	0.00008	0.34	0.02459	0.02480	5
7	0.02469	0.00009	0.37	0.02453	0.02476	5
6	0.02457	0.00007	0.29	0.02448	0.02463	5
5	0.02450	0.00001	0.06	0.02448	0.02451	5
4	0.02440	0.00007	0.29	0.02430	0.02450	5
3	0.02441	0.00008	0.35	0.02432	0.02455	5
2	0.02421	0.00013	0.53	0.02404	0.02434	5
1	0.02529	0.00021	0.82	0.02506	0.02548	5

Der gleiche Trend wie beim Hexan...Oktan bestätigt sich auch beim Cyclohexan. Kleinere  $\Delta_{\max}(1)$  Werte haben höhere RMSG-Werte und niedriger RMSDSN-Werte zur Folge. Für  $\Delta_{\max}(1)=1$  gibt es einen Ausreißer mit allseitig schlechten Qualitätsparametern. Mit Hinblick auf die Streuung der RMSG-Werte kann in dieser Trainingsreihe eine maximale Nachbarschaft von 4 Neuronen (40%) zu Start des Trainings als untere Grenzen angesehen werden.

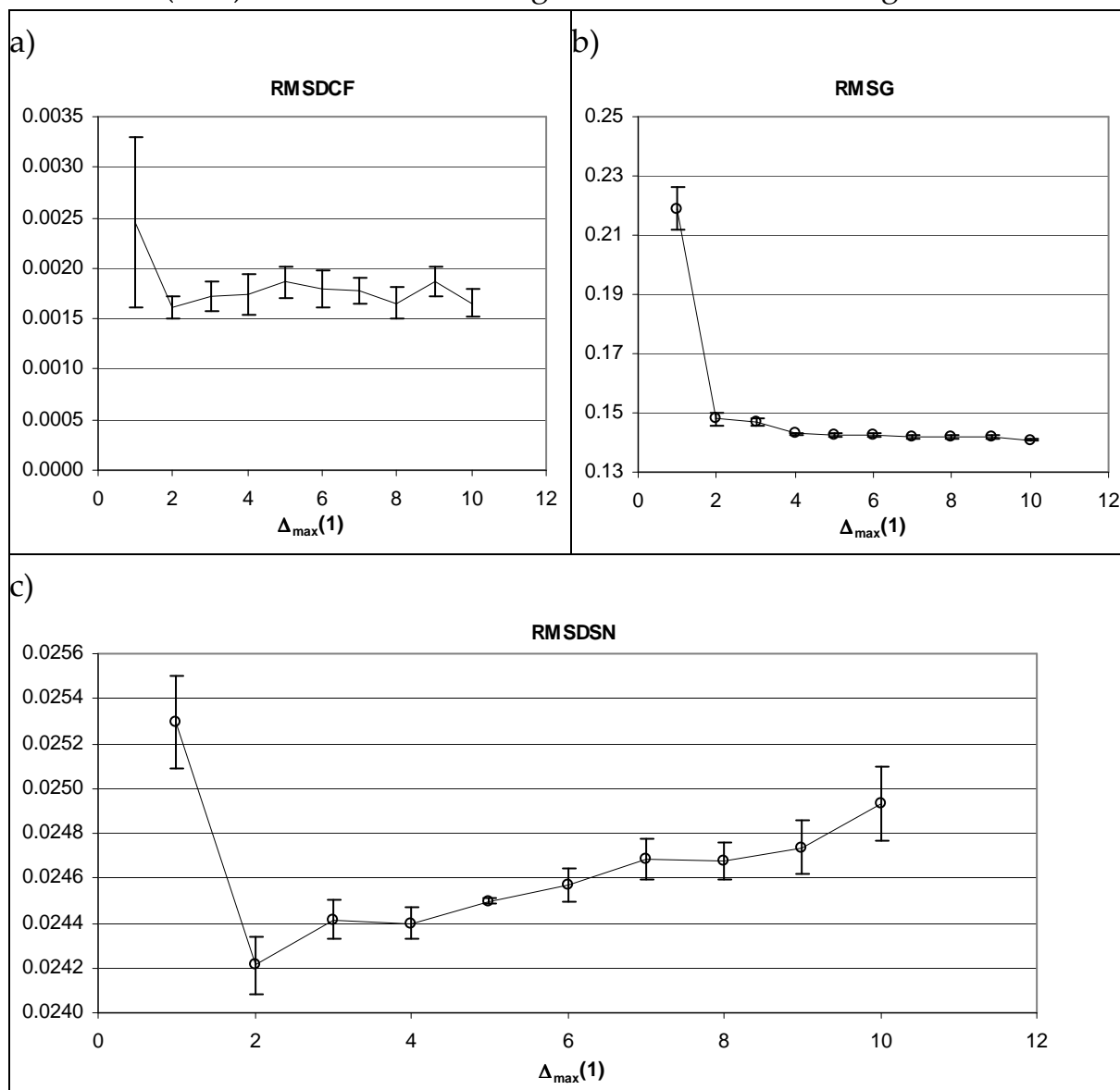


Abbildung 135: Variation von  $\Delta_{\max}(1)$ , a) RMSDCF, b) RMSG, c) RMSDSN



### 2.8.6.1.4 Variation von $\eta(\xi)$

Tabelle 64: Variation der Zeitabhängigkeit des Adaptionfaktors,  $\eta(\xi)$  nach Gleichung (2.17)

$\eta(\xi)$ (2.17)	$\emptyset$ RMSDCF	$\sigma$	relerr%	min	max	ndata
a) linear	0.00177	0.00019	10.8	0.00147	0.00194	5
b) quadratisch	0.00200	0.00006	3.0	0.00192	0.00206	5
$\eta(\xi)$ (2.17)	$\emptyset$ RMSG	$\sigma$	relerr%	min	max	ndata
a) linear	0.14164	0.000278	0.2	0.14143	0.14202	5
b) quadratisch	0.13098	0.000522	0.4	0.13013	0.13138	5
$\eta(\xi)$ (2.17)	$\emptyset$ RMSDSN	$\sigma$	relerr%	min	max	ndata
a) linear	0.02475	0.000142	0.6	0.02452	0.02488	5
b) quadratisch	0.03012	0.000097	0.3	0.03005	0.03024	5

Die grafische Auftragung obiger Tabelle in Abbildung 136 zeigt den gleichen Trend, welcher bereits vom Hexan bis hin zum Oktan beobachtet wird.

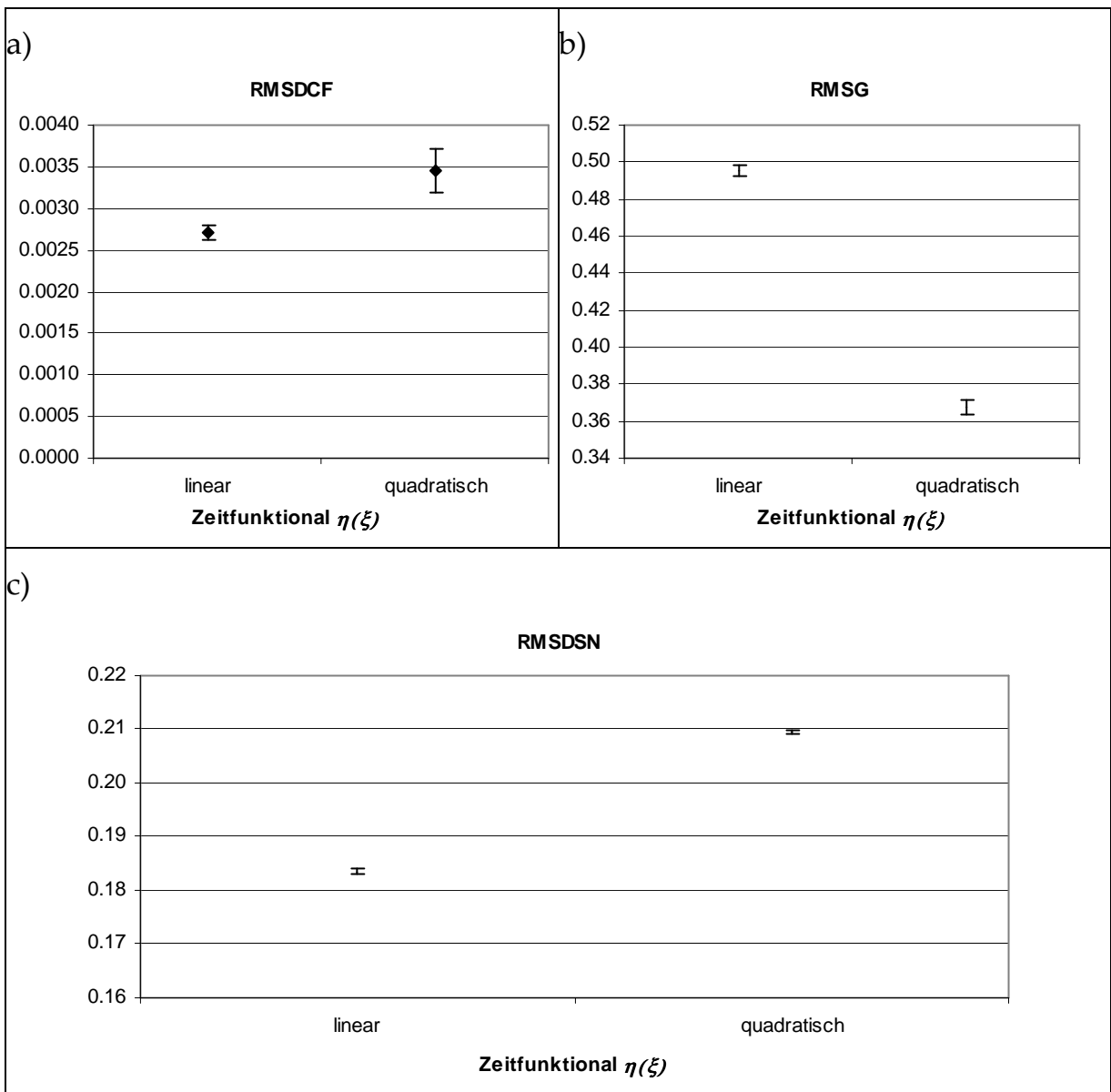


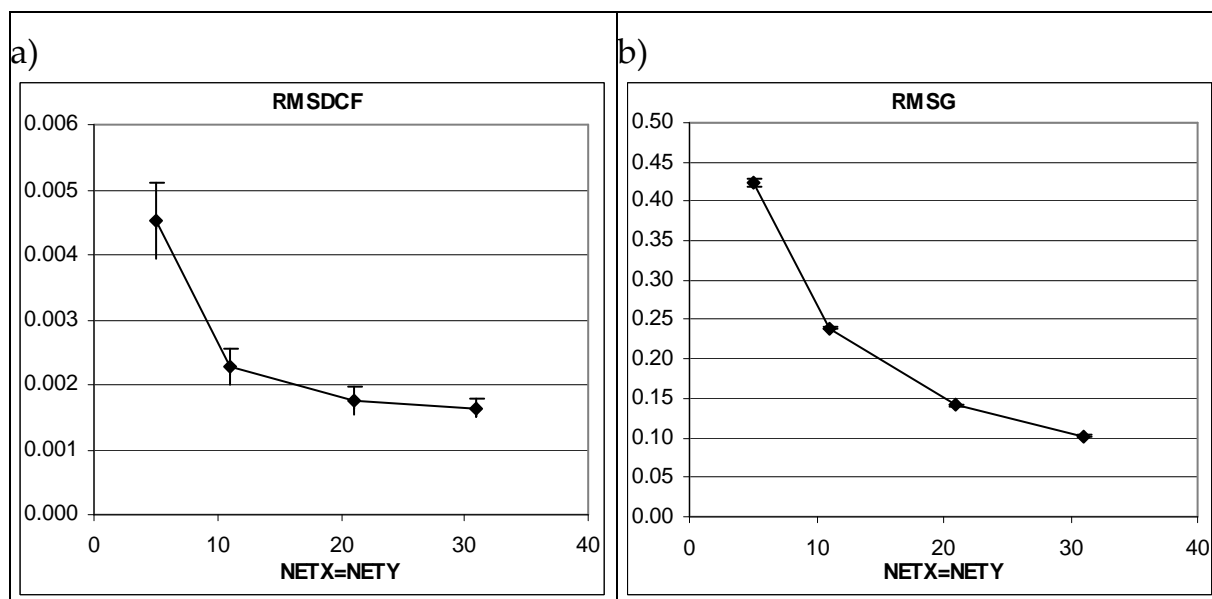
Abbildung 136: Variation von  $\eta(\xi)$  nach Gleichung (2.17)

### 2.8.6.1.5 Variation von NETX, NETY

Tabelle 65: Variation der Netzgröße NETX=NETY

NETX=NETY	Ø RMSDCF	$\sigma$	relerr%	min	max	ndata
31	0.00164	0.00014	8.63	0.00142	0.00175	5
21	0.00175	0.00022	12.30	0.00149	0.00205	5
11	0.00227	0.00028	12.41	0.00178	0.00247	5
5	0.00453	0.00059	13.00	0.00365	0.00526	5
NETX=NETY	Ø RMSG	$\sigma$	relerr%	min	max	ndata
31	0.10269	0.00065	0.63	0.10167	0.10340	5
21	0.14106	0.00057	0.40	0.14025	0.14163	5
11	0.23935	0.00145	0.61	0.23768	0.24088	5
5	0.42445	0.00470	1.11	0.41928	0.43035	5
NETX=NETY	Ø RMSDSN	$\sigma$	relerr%	min	max	ndata
31	0.01680	0.00009	0.53	0.01670	0.01690	5
21	0.02490	0.00016	0.63	0.02463	0.02502	5
11	0.04442	0.00008	0.19	0.04432	0.04453	5
5	0.08847	0.00057	0.64	0.08765	0.08911	5

In Abbildung 137 ist die grafische Auftragung der Tabellenwerte aus Tabelle 65 zu finden. Größere Kohonennetze ergeben bessere Projektionsqualitäten bei allen Kriterien.



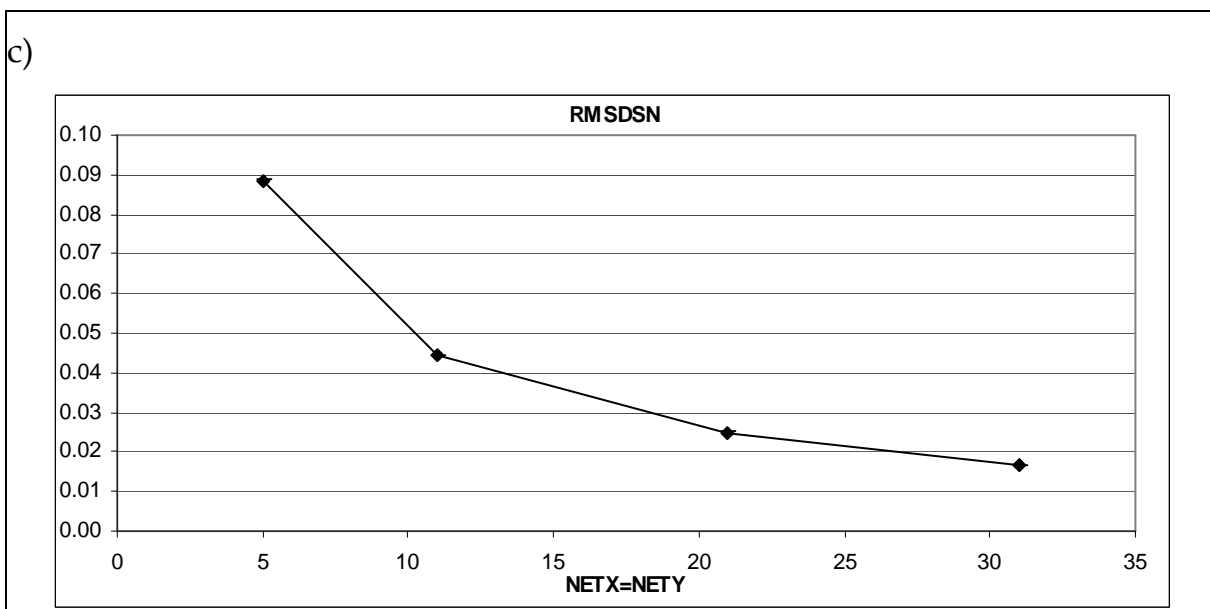


Abbildung 137: Variation der Netzgröße NETX=NETY

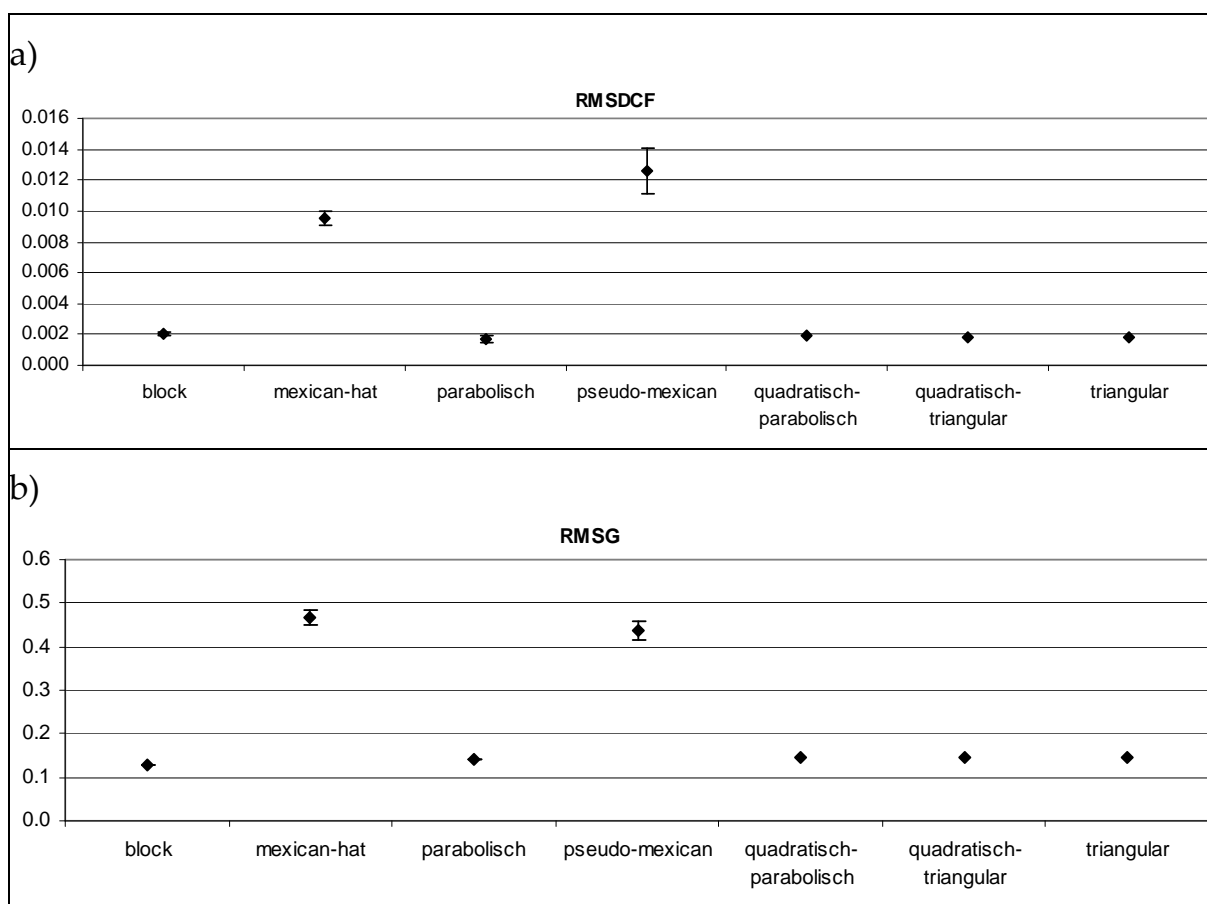
### 2.8.6.1.6 Variation der Abstandsfunktion $\vartheta(\zeta)$

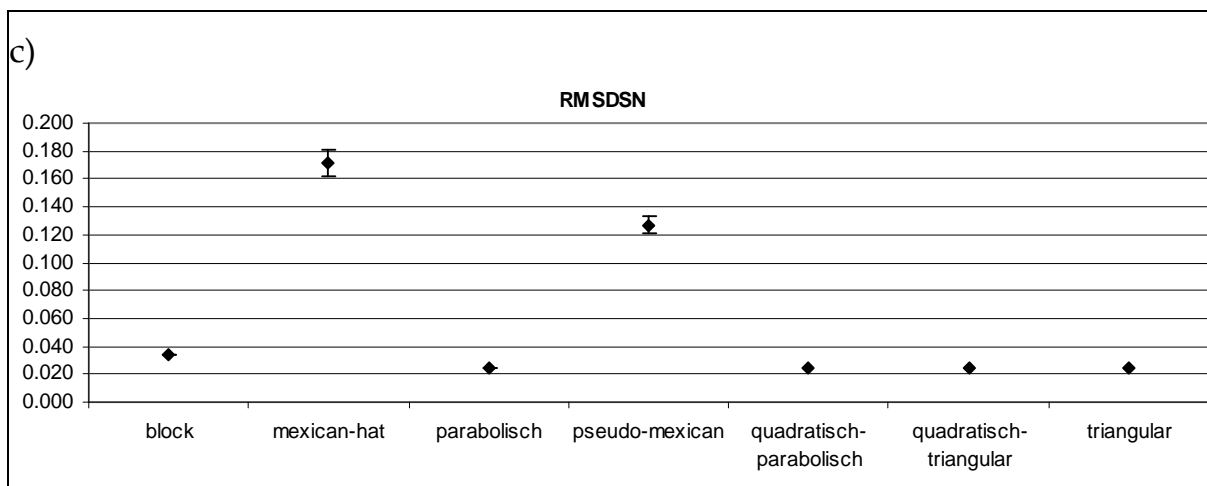
Tabelle 66: Variation der Abstandsfunktion  $\vartheta(\zeta)$  nach Gleichung (2.18)

$\vartheta(\zeta)$	$\emptyset$ RMSDCF	$\sigma$	relerr%	min	max	ndata
<i>block</i>	0.00201	0.00010	4.88	0.00191	0.00216	5
<i>Mexican hat</i>	0.00956	0.00045	4.68	0.00897	0.00997	5
parabolisch	0.00170	0.00022	12.72	0.00139	0.00190	5
<i>pseudo Mexican</i>	0.01257	0.00146	11.64	0.01092	0.01489	5
quadratisch- parabolisch	0.00192	0.00016	8.33	0.00174	0.00209	5
quadratisch- triangular	0.00186	0.00019	10.41	0.00167	0.00207	5
triangular	0.00178	0.00024	13.46	0.00155	0.00216	5
$\vartheta(\zeta)$	$\emptyset$ RMSG	$\sigma$	relerr%	min	max	ndata
<i>block</i>	0.13024	0.00040	0.30	0.12990	0.13090	5
<i>Mexican hat</i>	0.46631	0.01682	3.61	0.43941	0.48562	5
parabolisch	0.14124	0.00050	0.35	0.14076	0.14198	5
<i>pseudo Mexican</i>	0.43886	0.02156	4.91	0.40507	0.46336	5
quadratisch- parabolisch	0.14637	0.00040	0.27	0.14572	0.14664	5
quadratisch- triangular	0.14678	0.00063	0.43	0.14586	0.14738	5
triangular	0.14626	0.00028	0.19	0.14585	0.14652	5

$\mathfrak{S}(\zeta)$	$\emptyset$ RMSDSN	$\sigma$	relerr%	min	max	ndata
<i>block</i>	0.03400	0.00014	0.42	0.03383	0.03420	5
<i>Mexican hat</i>	0.17161	0.00971	5.66	0.15423	0.17595	5
parabolisch	0.02485	0.00012	0.49	0.02465	0.02497	5
<i>pseudo Mexican</i>	0.12683	0.00621	4.90	0.12087	0.13336	5
quadratisch- parabolisch	0.02426	0.00010	0.42	0.02412	0.02436	5
quadratisch- triangular	0.02410	0.00009	0.39	0.02396	0.02419	5
triangular	0.02416	0.00005	0.22	0.02407	0.02420	5

Der Trend aus den vorherigen Evaluierungen bestätigt sich, die Besprechung wird an dieser Stelle nicht wiederholt.





**Abbildung 138: Variation der Abstandsfunktion  $\vartheta(\zeta)$**

Auch für das Cyclohexanmolekül werden exemplarische Netzkarten für alle Abstandsfunktionen erzeugt und in Abbildung 139 wiedergegeben. Energiewerte, welche höher als die verwendete Skala sind, werden als graue Flächen wiedergegeben. Deutlich erkennbar sind die großen grauen Farbflächen in den *Mexican hat*-artigen Netzkarten. Dies ergibt sich aus den negativen Trainingsbereichen am Rande der *Mexican-hat*-Funktionen. Mit immer kleinerer trainierter Nachbarschaft kontrahiert der Konformationsbereich um die nahe Nachbarschaft eines zentralen Prototypneurons, während die entfernte Nachbarschaft mit invertierten Werten trainiert wird. Im Falle der fixierten Ringdiederwinkel im Cyclohexan bedeutet dies sterisch unmögliche *restraints* für MINIX1. Das Resultat sind extrem hohe Energiewerte durch das Verletzen dieser *restraints*. Die tatsächlichen Konformationen werden auf diese Weise klar voneinander durch „unmögliche“ Bereiche abgetrennt. Dieses Verhalten ist für andere Anwendungen durchaus erwünscht. Die Probleme der korrekten Identifizierung von im Trainingssatz tatsächlich vorhandenen und durch invertiertes Training „erzeugten“ Konformationen sind bei der Konformationsanalyse mittels Kohonenkarten jedoch zu groß.

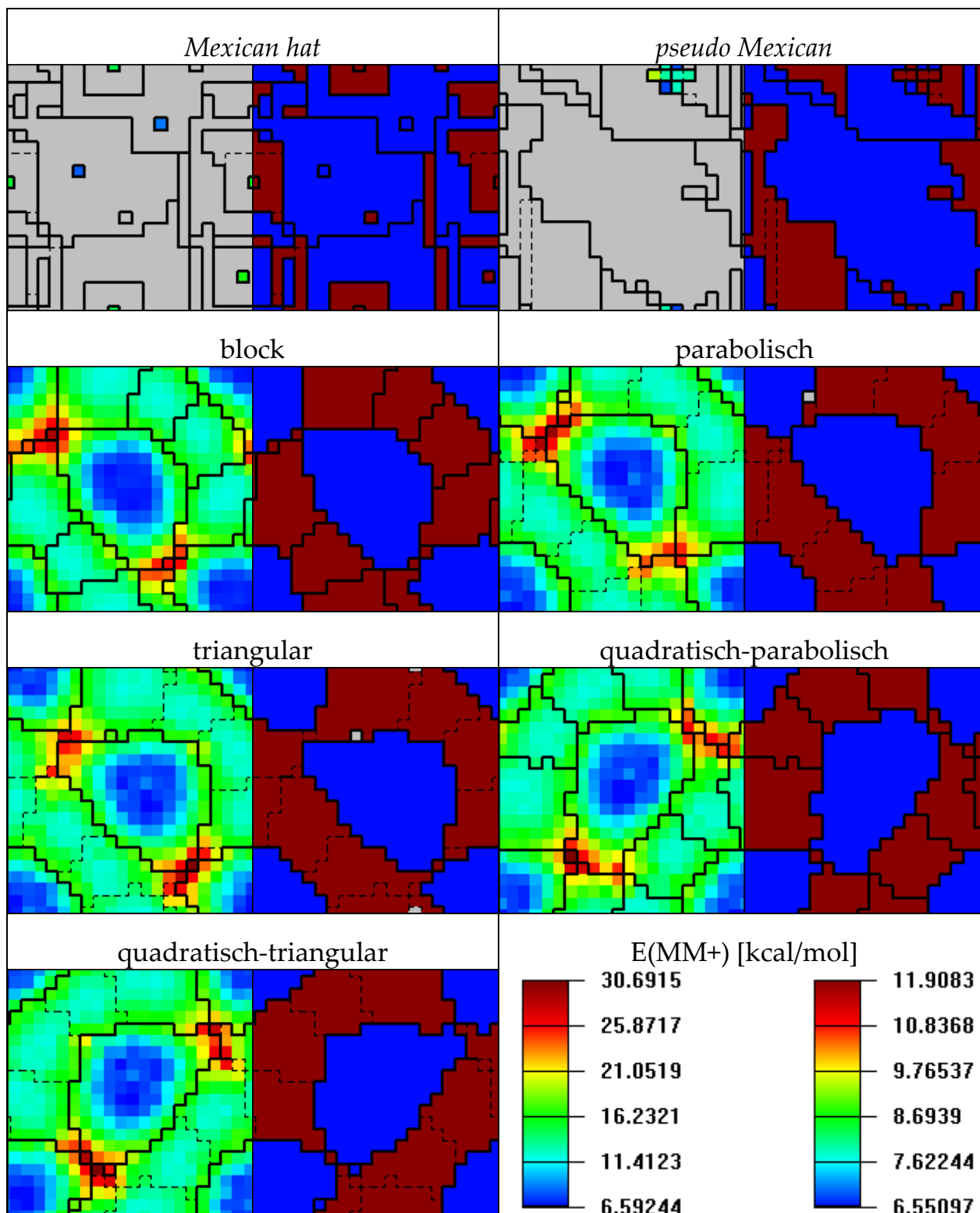


Abbildung 139: Beispielnetzarten Gridhyperfläche – Variation der Abstandsfunktion, Oktan, links partiell optimiert unter Verwendung der Netzrestraints nach COLORIX1, rechts vollständig optimiert nach COLORIX2

### 2.8.6.1.7 Fazit

Die Qualität der erhaltenen Cyclohexan-Netzarten ist wie beim Pentan ausgezeichnet, was eindrucksvoll z.B. in Abbildung 131 zum Ausdruck kommt. Auch wenn an dieser Stelle sechs interne Koordinaten trainiert

werden, liegt die wahre Dimensionalität der trainierten Hyperfläche eher zwischen zwei und drei, was die Grundlage der bereits erwähnten *ring-puckering*-Koordinaten darstellt.<sup>[112],[113]</sup>

Es kann damit gezeigt werden, dass Kohonennetze in der Lage sind, auch komplexe inhärente Dimensionsreduzierungen in Datensätzen zu erkennen und zu nutzen. Die absoluten RMSDCF- und RMSDSN-Werte bewegen sich im gleichen Größenordnungsbereich wie beim Pentan. Der RMSG-Wert befindet sich größenordnungsmäßig in einem vollkommen anderen Wertebereich und kann mit Sicherheit nur relativ verwendet werden.

Es wurden folgende MTS Werte (minimale Trainingsschritte nach Kapitel 2.5.4) ermittelt:

- MTS(RMSG, 1%) > 3366000
- MTS(RMSDSN, 1%) > 3366000

### 2.8.6.2 MD Cyclohexan, Training der Diederwinkel

Für das Cyclohexanmolekül wird bei unterschiedlichen Temperaturen eine Gruppe Moleküldynamikrechnungen durchgeführt. (Kraftfeld HYPERCHEM MM+, in vacuo, 200000 Strukturen, Simulationsdauer 100000 ps, Aufruf von FIX\_ROTATION.EXE bei je 10 ps, *step size* 0.001 ps, *datafreq*=1/500 *data steps*<sup>-1</sup>). Damit werden neuronale Netze trainiert. [25x25 Neuronen; 100 *epochs*;  $\Delta_{\max}(1)=12$ ;  $f_0=0.4$ ; Abstandsfunktion parabolisch, Training der 6 CC-Ringdiederwinkel]. Die erhaltenen Qualitätsparameter sind in Tabelle 67 aufgelistet. Eine Netzkarte mit namentlich identifizierten und den Clustern der Oberfläche zugeordneten Konformationen ist bereits in Abbildung 131 aufgeführt.

**Tabelle 67:** Cyclohexan MD-Simulationen bei unterschiedlichen Temperaturen, Training der Ringdiederwinkel

T [K]	Ø RMSDCF	$\sigma$	relerr%	min	max	ndata
1000	0.00288	0.00019	6.46	0.00265	0.00306	5
900	0.00338	0.00023	6.70	0.00309	0.00372	5
800	0.00362	0.00021	5.67	0.00338	0.00382	5
700	0.00368	0.00025	6.82	0.00331	0.00390	5
600	0.00320	0.00008	2.57	0.00309	0.00332	5
500	0.00250	0.00027	10.96	0.00213	0.00284	5

T [K]	Ø RMSG	$\sigma$	relerr%	min	max	ndata
1000	0.14757	0.00238	1.61	0.14486	0.14936	5
900	0.15245	0.00382	2.51	0.14933	0.15697	5
800	0.16288	0.00648	3.98	0.15597	0.16994	5
700	0.17629	0.00481	2.73	0.16979	0.17998	5
600	0.18980	0.00354	1.86	0.18724	0.19604	5
500	0.20367	0.00263	1.29	0.20072	0.20655	5
T [K]	Ø RMSDSN	$\sigma$	relerr%	min	max	ndata
1000	0.03319	0.00002	0.07	0.03315	0.03320	5
900	0.03143	0.00002	0.06	0.03142	0.03146	5
800	0.02940	0.00002	0.08	0.02937	0.02942	5
700	0.02693	0.00003	0.11	0.02691	0.02698	5
600	0.02407	0.00004	0.16	0.02400	0.02410	5
500	0.02183	0.00003	0.12	0.02181	0.02187	5

Die kontinuierlich fallenden RMSG- und steigenden RMSDSN-Werte sind wie beim Pentan Indikator für eine Projektion ohne große Projektionsverluste.

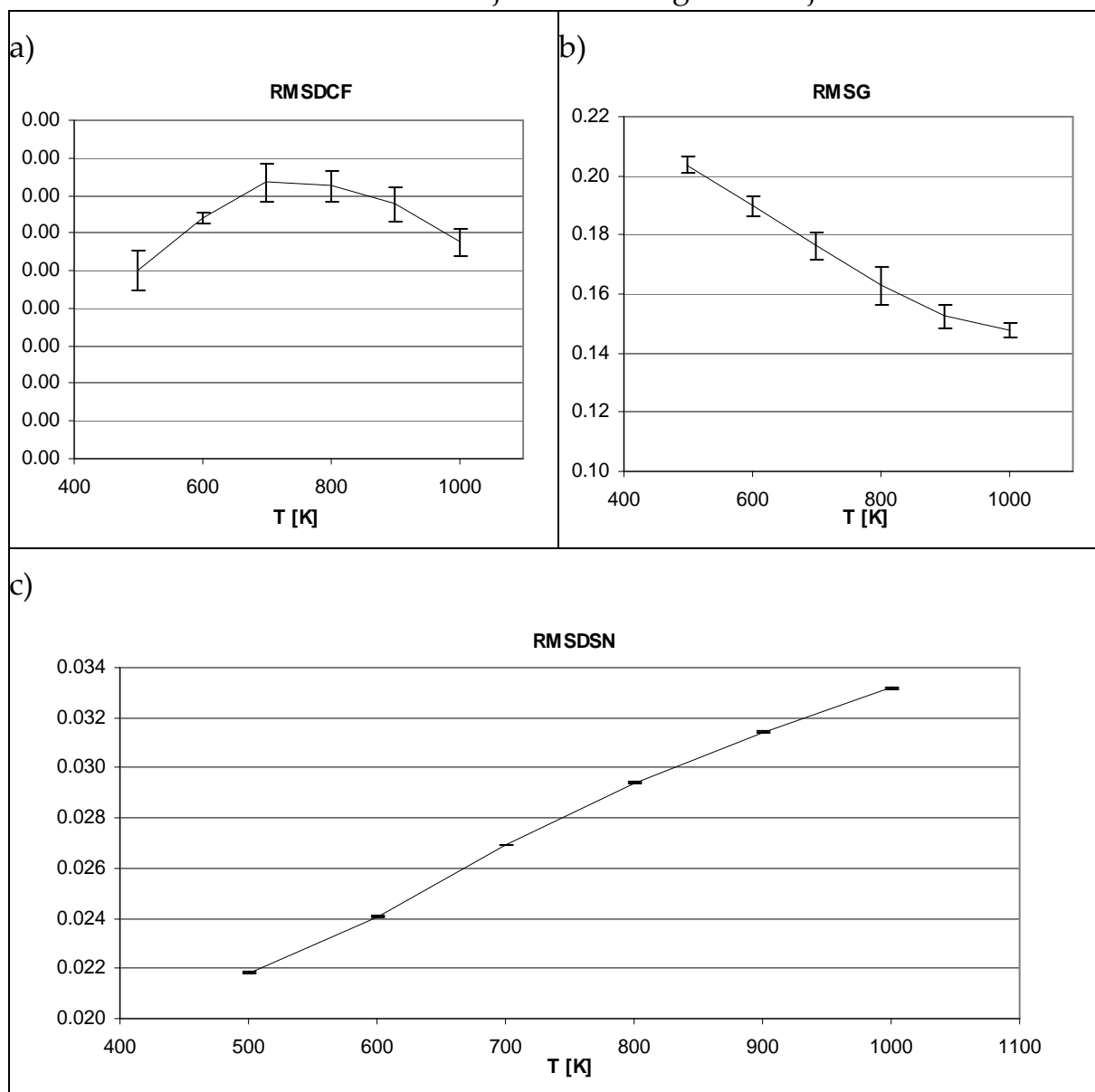


Abbildung 140: Variation der Temperatur – Cyclohexan MD, Training der drei CC-Diederwinkel



Exemplarische Netzkarten nach Durchführung des MINIX1- und MINIX2-Moduls sind in Abbildung 141 wiedergegeben. Das Maximum im RMSDCF-Graphen in Abbildung 140 zeigt einen Wechsel der Netztopologie zwischen 700K und 800K an, was dem Zusammenwachsen der *twist* Minima zu einem Band entspricht.

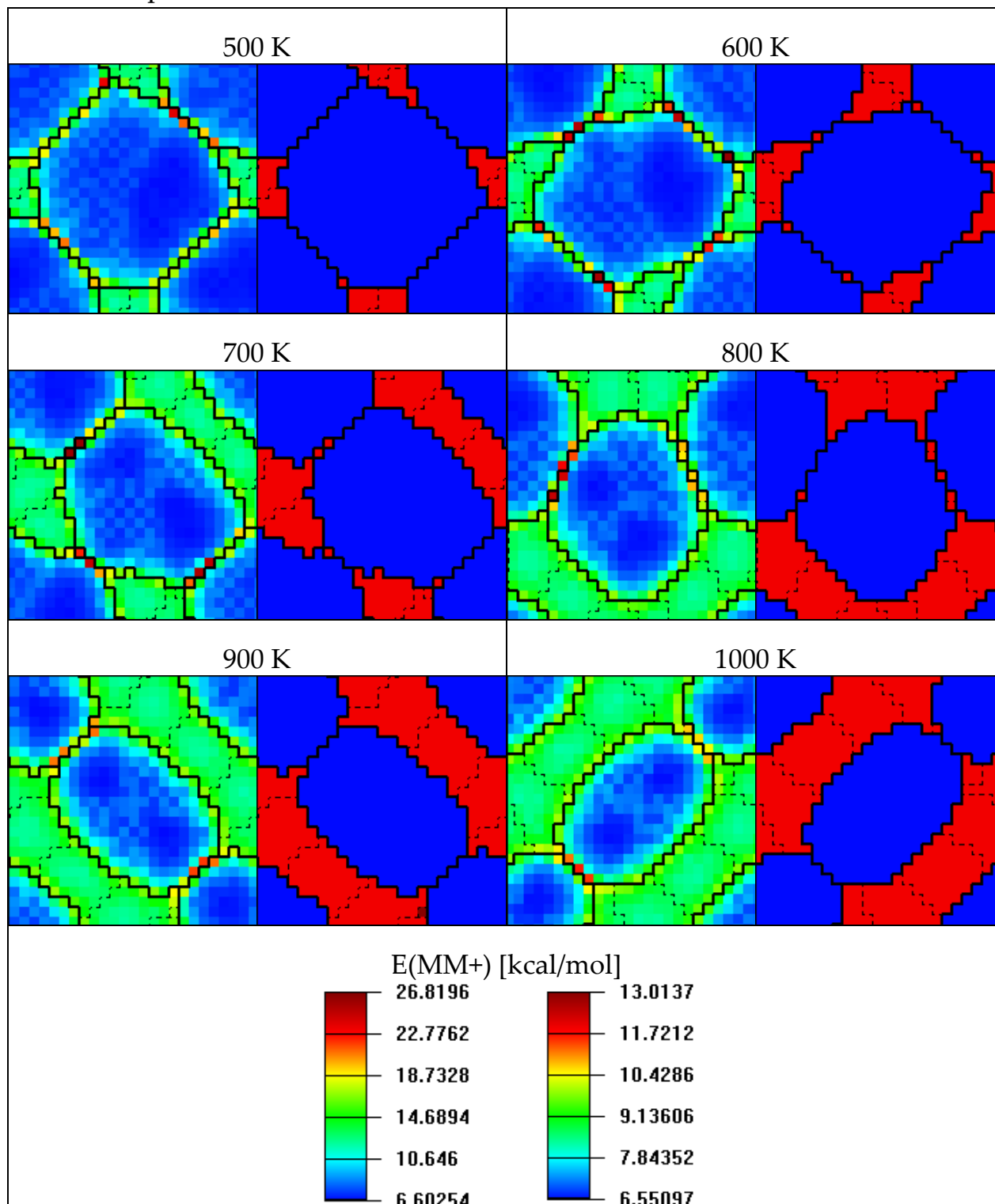


Abbildung 141: Beispielnetzarten Gridhyperfläche – Variation der Temperatur – Cyclohexan MD, Training der 6 CC-Ringdiederwinkel, links partiell optimiert unter Verwendung der Netzrestraints nach COLORIX1, rechts vollständig optimiert nach COLORIX2

### 2.8.6.3 MD Cyclohexan, Training 26 zufälliger Distanzen

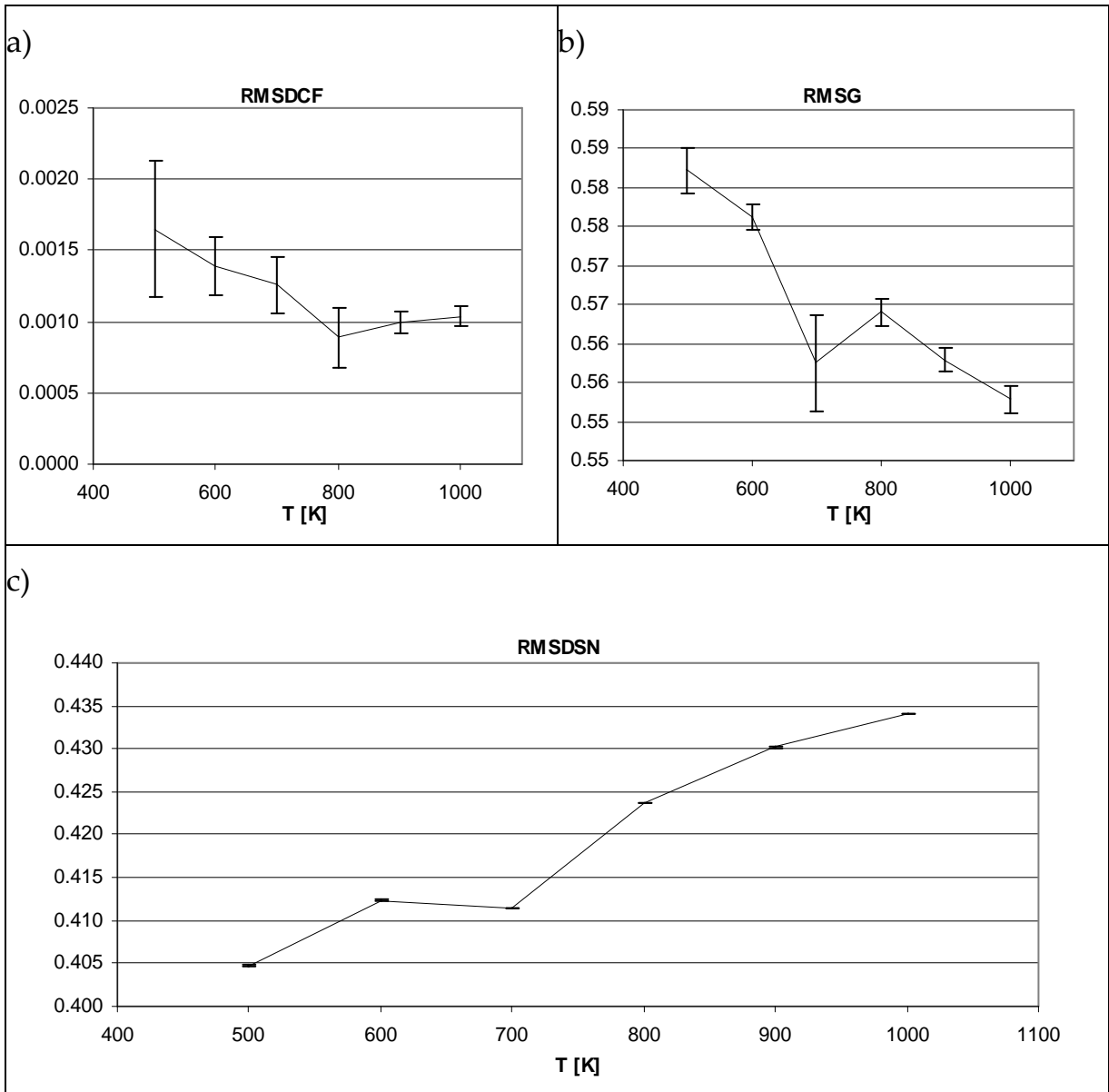
Mit den gleichen Moleküldynamiken wie im vorherigen Kapitel werden Kohonennetze mit einem Satz von sechsundzwanzig zufällig ausgewählten C-C, C-H und H-H Distanzen trainiert.

Die statistischen Daten des Netztrainings sind in Tabelle 68 aufgeführt. [25x25 Neuronen; 20 *epochs*;  $\Delta_{\max}(1)=12$ ;  $f_0=0.4$ ; Abstandsfunktion parabolisch, Training 26 zufällig ausgewählter Distanzen].

**Tabelle 68: Cyclohexan MD-Simulationen bei unterschiedlichen Temperaturen, Training 26 zufällig ausgewählter Distanzen**

T [K]	Ø RMSDCF	$\sigma$	relerr%	min	max	ndata
1000	0.00104	0.00007	6.96	0.00098	0.00116	5
900	0.00100	0.00008	7.79	0.00088	0.00108	5
800	0.00089	0.00022	24.28	0.00055	0.00114	5
700	0.00126	0.00020	15.68	0.00102	0.00146	5
600	0.00139	0.00020	14.38	0.00115	0.00161	5
500	0.00165	0.00048	28.90	0.00090	0.00221	5
T [K]	Ø RMSG	$\sigma$	relerr%	min	max	ndata
1000	0.55286	0.00175	0.32	0.55022	0.55468	5
900	0.55793	0.00159	0.29	0.55566	0.56004	5
800	0.56401	0.00175	0.31	0.56246	0.56660	5
700	0.55750	0.00622	1.12	0.55196	0.56824	5
600	0.57620	0.00167	0.29	0.57361	0.57774	5
500	0.58222	0.00288	0.49	0.57770	0.58541	5
T [K]	Ø RMSDSN	$\sigma$	relerr%	min	max	ndata
1000	0.43401	0.00003	0.01	0.43398	0.43405	5
900	0.43022	0.00009	0.02	0.43008	0.43031	5
800	0.42374	0.00006	0.01	0.42369	0.42382	5
700	0.41139	0.00005	0.01	0.41133	0.41146	5
600	0.41238	0.00009	0.02	0.41229	0.41248	5
500	0.40476	0.00005	0.01	0.40469	0.40483	5

Wie bei den Molekülen zuvor zeigen die in Abbildung 142 wiedergegeben Kurven mit steigender Temperatur fallende RMSG- und steigende RMSDSN-Werte.



**Abbildung 142: Variation der Temperatur – Oktan MD, Training 26 zufällig ausgewählter Distanzen**

Der Knick bei 700 K in den Kurven der Qualitätsparameter koinzidiert mit einem Wechsel der Netztopologie von bandförmigen zu kreisförmigen *chair*-Clustern. Die Netzoberflächen wirken insgesamt verrauschter und unsymmetrischer als die Diedernezzkarten. Damit kann für kleine Ringsysteme eine bessere Eignung der Diederwinkel konstatiert werden.

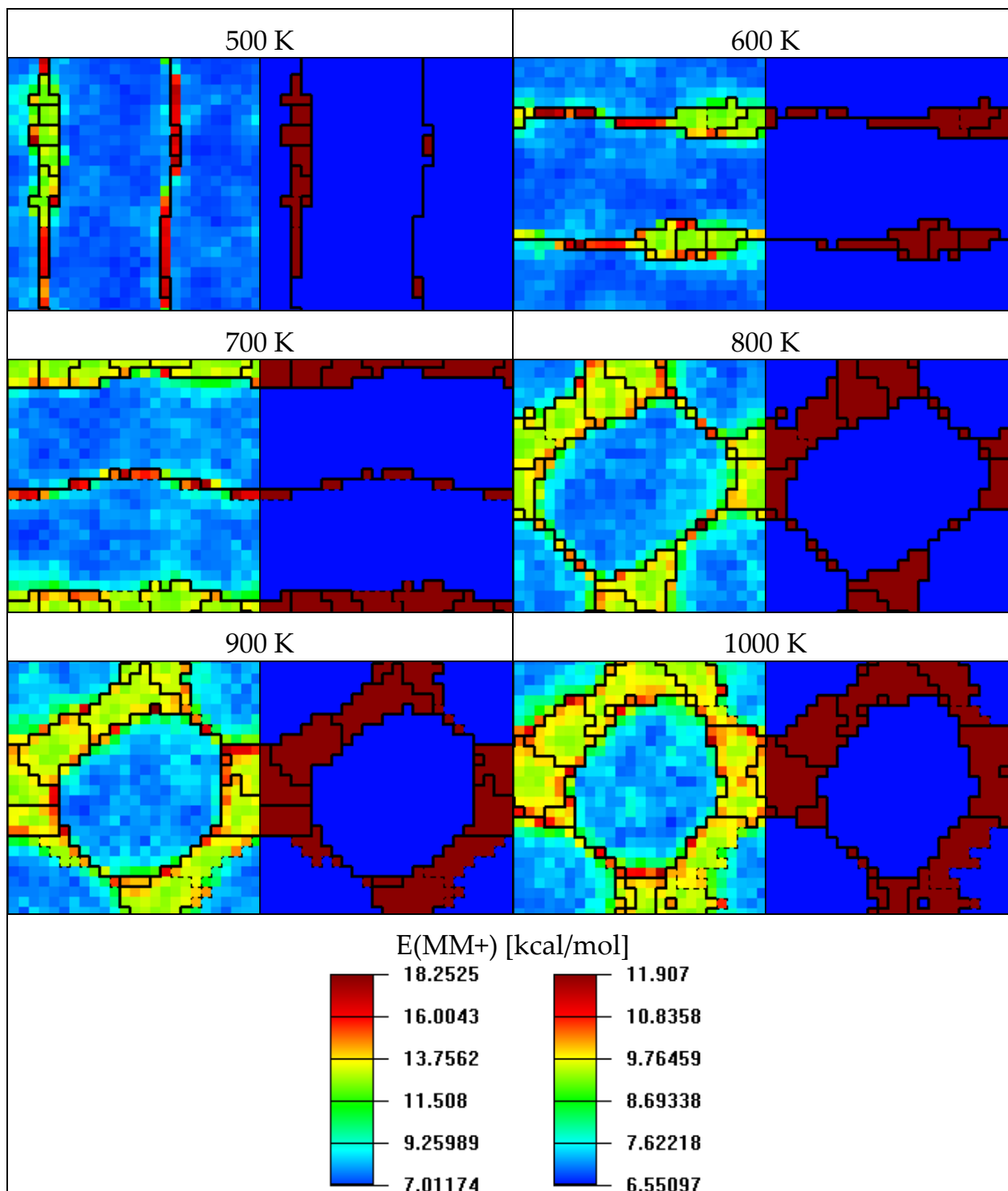


Abbildung 143: Beispielnetzkarten Gridhyperfläche – Variation der Temperatur – Cyclohexan MD, Training 26 zufällig ausgewählter Distanzen, links partiell optimiert unter Verwendung der Netzrestraints nach COLORIX1, rechts vollständig optimiert nach COLORIX2

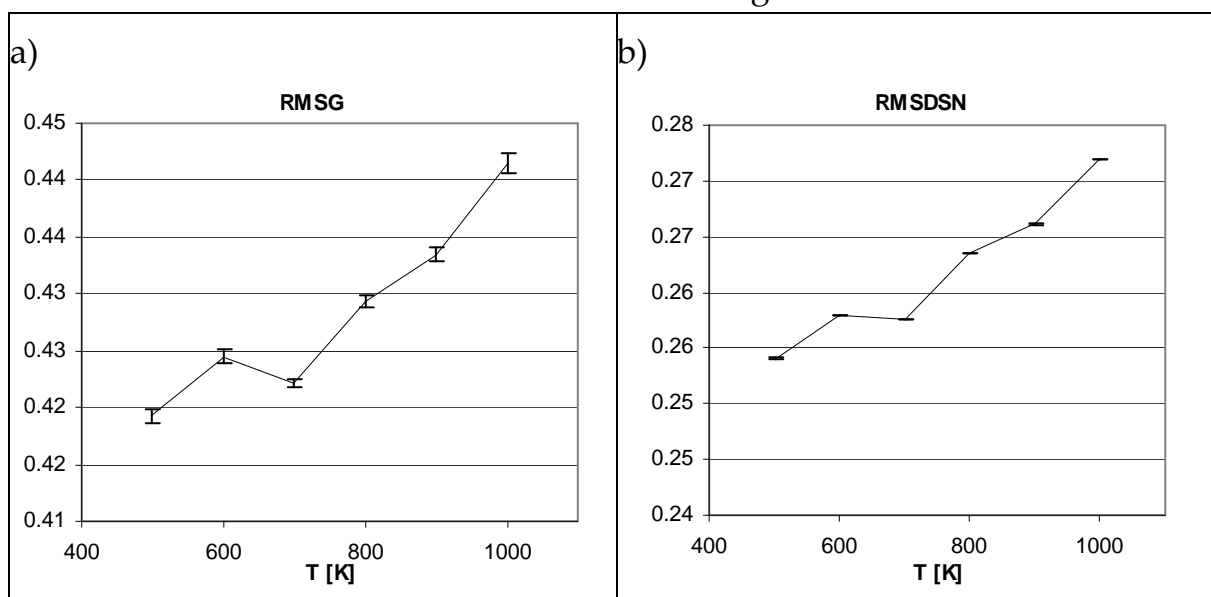
#### 2.8.6.4 MD Cyclohexan, Training der Schweratomdistanzen

Im Folgenden werden die 15 C-C Abstände des Rings als Eingabevektoren des Netztrainings verwendet. [25x25 Neuronen; 20 *epochs*;  $\Delta_{\max}(1)=12$ ;  $f_0=0.4$ ; Abstandsfunktion parabolisch, Training der CC-Distanzen], siehe Tabelle 69.

**Tabelle 69: Cyclohexan MD-Simulationen bei unterschiedlichen Temperaturen, Training der Schweratomdistanzmatrix**

T [K]	Ø RMSG	$\sigma$	relerr%	min	max	ndata
1000	0.44143	0.00085	0.19	0.44055	0.44242	5
900	0.43346	0.00056	0.13	0.43269	0.43400	5
800	0.42936	0.00060	0.14	0.42877	0.43024	5
700	0.42211	0.00035	0.08	0.42180	0.42267	5
600	0.42449	0.00057	0.13	0.42393	0.42528	5
500	0.41927	0.00061	0.14	0.41863	0.42021	5
T [K]	Ø RMSDSN	$\sigma$	relerr%	min	max	ndata
1000	0.27198	0.00003	0.01	0.27193	0.27201	5
900	0.26614	0.00003	0.01	0.26609	0.26618	5
800	0.26351	0.00003	0.01	0.26347	0.26354	5
700	0.25760	0.00003	0.01	0.25756	0.25763	5
600	0.25796	0.00007	0.03	0.25789	0.25808	5
500	0.25405	0.00004	0.01	0.25402	0.25411	5

Die Auswertung der Graphen aus Abbildung 144 erübrigt sich nach einem Blick auf die erhaltenen Netzkarten in Abbildung 145.

**Abbildung 144: Variation der Temperatur – Cyclohexan MD, Training der Schweratomdistanzmatrix**

Die *twist*-Minima liegen als einzelne Punkte quasistochastisch über die Netzoberfläche verteilt. Offensichtlich gehen die Varianzen der Abstände der Ringatome komplett im thermischen Rauschen der Simulation unter. Als einziges Ergebnis erhält man die komplette Nichteignung der Ringatomdistanzen als Eingabewerte des Netztrainings für kleine Ringsysteme.

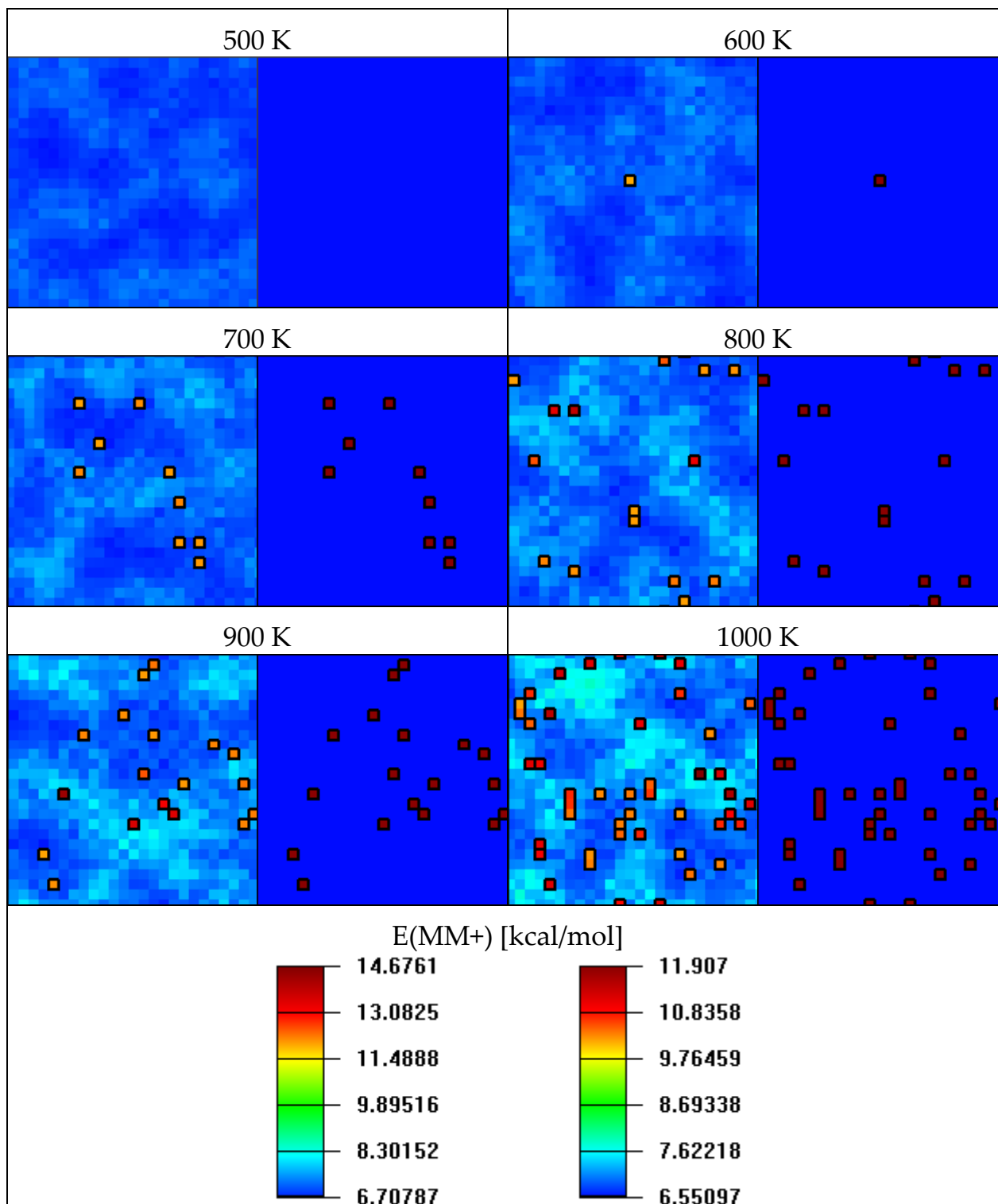


Abbildung 145: Beispielnetzarten Gridhyperfläche – Variation der Temperatur – Cyclohexan MD, Training der Schweratomdistanzmatrix, links partiell optimiert unter Verwendung der Netzrestraints nach COLORIX1, rechts vollständig optimiert nach COLORIX2

### 2.8.6.5 Trajektorien Cyclohexan

Wie bereits in Kapitel 2.8.1.4 beschrieben, kann der Hypersurface Explorer verwendet werden, um Konformationsdatensätze, meist MD-Daten, auf eine Hyperfläche zu projizieren. Für das Cyclohexanmolekül soll dies anhand einer in Kapitel 2.8.6.1.6 erhaltenen Netzkarte (triangular) demonstriert werden.

Dazu werden kurze Moleküldynamiken startend von einer *chair* Konformation generiert [MM+; 40 x 1 ps FIX\_ROTATION.EXE; *step size* 0.001 ps; *datarate*=1/20; - > 2000 Strukturen]. Die Trajektorien werden in Abbildung 146 auf die deterministische Hyperfläche der Netzkarte projiziert.

Der abgeschrittene Konformationsraum ist bei 1000K ganz klar am größten. Des Weiteren werden die Bewegungen auf der Hyperfläche bei höheren Temperaturen in größeren Schritten vollzogen. Bei 800 K und 1000 K überwindet das Molekül mehrfach die hohe Barriere von *chair*- nach *twist*-Konformation. Bei 600 K gelingt dieser Übergang zweimal. Bei 400 K ist das Molekül nach Überwinden des energetisch hoch liegenden *half-chair*-Übergangszustands in den *twist*-Konformationen „gefangen“, deren niedrige *boat*-Übergangszustände problemlos überwunden werden. Bei 300 K schließlich bleibt das Molekül komplett in der *chair*-Startkonformation.

Eine thermodynamische Auswertung zur Bestimmung der Aktivierungsbarriere ist jedoch nicht möglich, da die Simulation als Einzelmolekül im Temperaturbad, sowie die explizite Dämpfung der Rotations- und Translationsfreiheitsgrade mit Hilfe des Programms FIX\_ROTATION, siehe Kapitel 5.3.8, das thermodynamische Ensemble verfälschen.

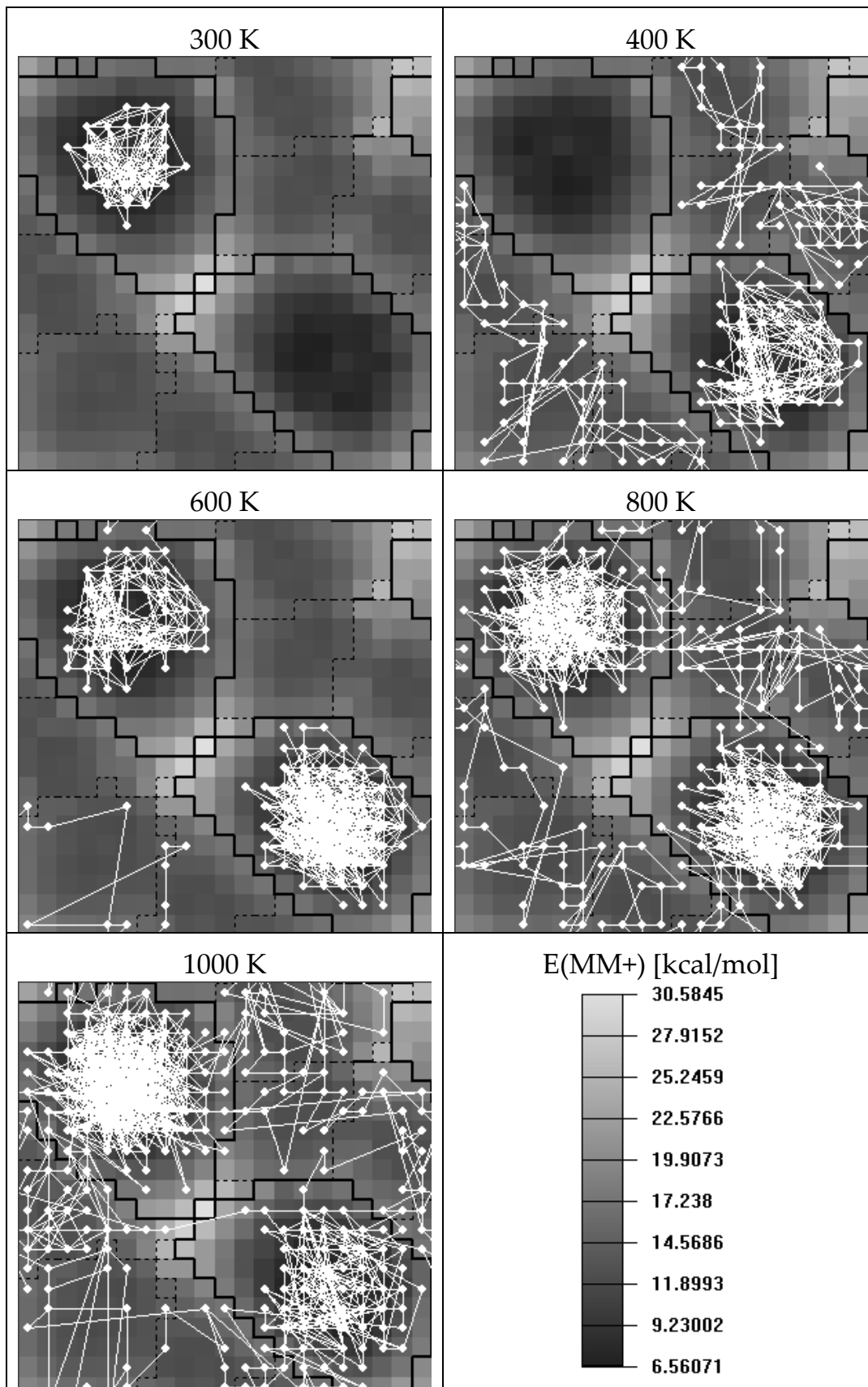


Abbildung 146: Cyclohexan – 40 ps MD-Trajektorien projiziert auf die deterministische Hyperfläche



### 2.8.7 *Cis-Decalin*

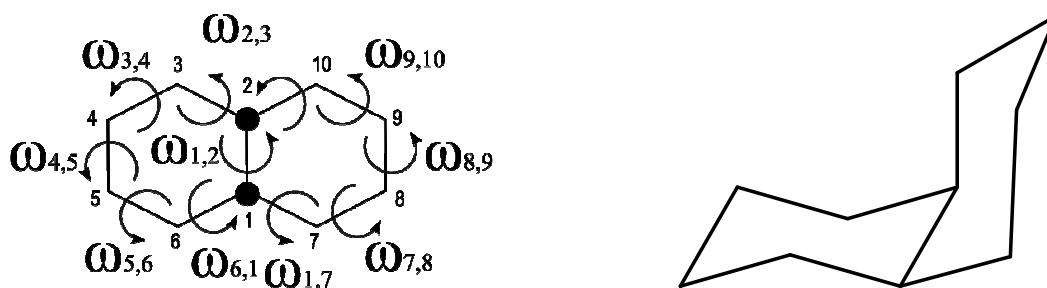


Abbildung 147: CC-Diederwinkel im *cis-Decalin*

Decalin existiert in zwei Konfigurationen, dem *cis*- und dem *trans*-Isomer. An dieser Stelle wird ausschließlich das *cis*-Decalin untersucht. Die Rigidität der annellierten Ringe ist ein Motiv, das in vielen Biomolekülen, z.B. Hormonen (dort allerdings eher das *trans*-Isomer) wiederkehrt. Während die Anzahl der Atome größer ist als z.B. im Oktan, sind die Freiheitsgrade wesentlich stärker eingeschränkt, so dass aussagekräftigere Netzkarten erwartet werden. Die Erstellung einer deterministischen Hyperfläche durch Variation der Diederwinkel wird aufgrund der großen Zahl Diederwinkel und der sterischen Beschränkungen als zu aufwändig verworfen. Es werden eine Reihe Moleküldynamiken bei unterschiedlichen Temperaturen durchgeführt. (Kraftfeld HYPERCHEM MM+, in vacuo, 200000 Strukturen, Simulationsdauer 200000 ps, Aufruf von FIX\_ROTATION.EXE bei je 20 ps, *step size* 0.001 ps, *datafreq*=1/1000 *data steps*<sup>-1</sup>).

#### 2.8.7.1 MD *cis-Decalin*, Training der Diederwinkel

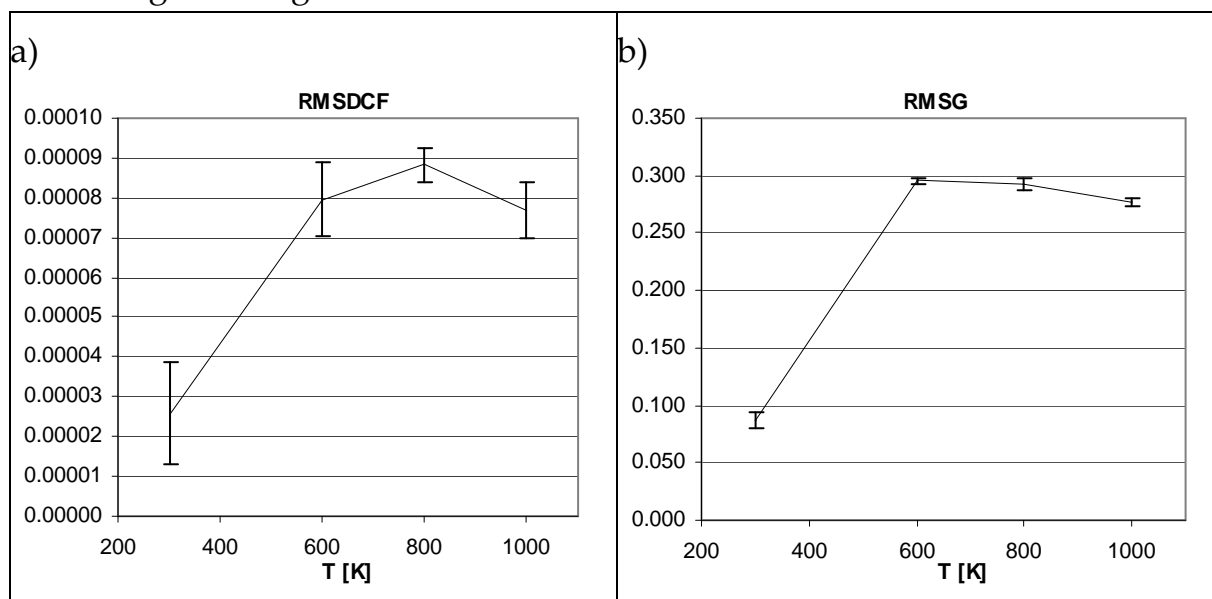
Mit den in Abbildung 147 angegebenen Diederwinkeln, extrahiert aus obigen Moleküldynamiken, werden Kohonennetze trainiert [25x25 Neuronen; 100 *epochs*;  $\Delta_{\max}(1)=12$ ;  $f_0=0.4$ ; Abstandsfunktion parabolisch, Training der 11 CC-Ringdiederwinkel]. Die RMSDCF-Werte werden wiederum mit Hilfe des Programms CLASSIFY\_SNP\_CONFORMATIONS.EXE berechnet. Die erhaltenen Qualitätsparameter sind in Tabelle 70 aufgeführt.

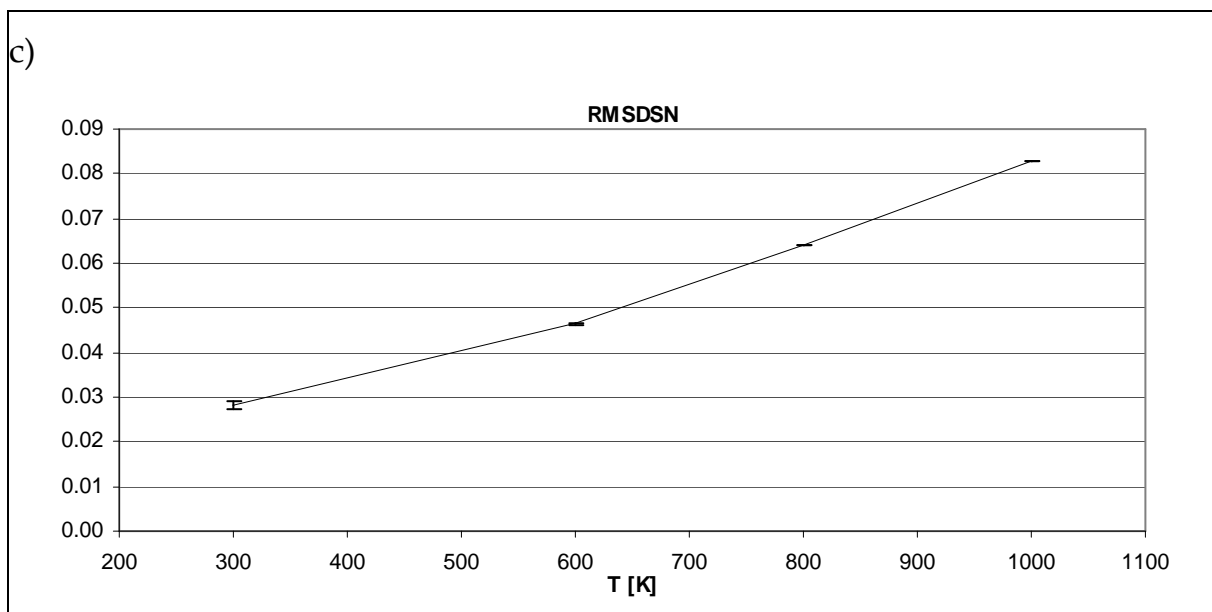
Tabelle 70: Decalin MD-Simulationen bei unterschiedlichen Temperaturen, Training der 11 Ringdiederwinkel

T [K]	Ø RMSDCF	$\sigma$	relerr%	min	max	ndata
1000	7.675E-05	6.999E-06	9.12	6.955E-05	8.660E-05	5
800	8.834E-05	4.374E-06	4.95	8.228E-05	9.363E-05	5
600	7.957E-05	9.334E-06	11.73	6.407E-05	8.700E-05	5
300	2.586E-05	1.285E-05	49.68	1.004E-05	3.934E-05	5

T [K]	Ø RMSG	$\sigma$	relerr%	min	max	ndata
1000	0.27619	0.00363	1.31	0.27171	0.27953	5
800	0.29247	0.00529	1.81	0.28808	0.29980	5
600	0.29574	0.00264	0.89	0.29236	0.29893	5
300	0.08719	0.00703	8.06	0.07877	0.09259	5
T [K]	Ø RMSDSN	$\sigma$	relerr%	min	max	ndata
1000	0.08293	0.00010	0.13	0.08277	0.08303	5
800	0.06411	0.00009	0.14	0.06401	0.06424	5
600	0.04642	0.00026	0.55	0.04599	0.04667	5
300	0.02809	0.00097	3.44	0.02736	0.02917	5

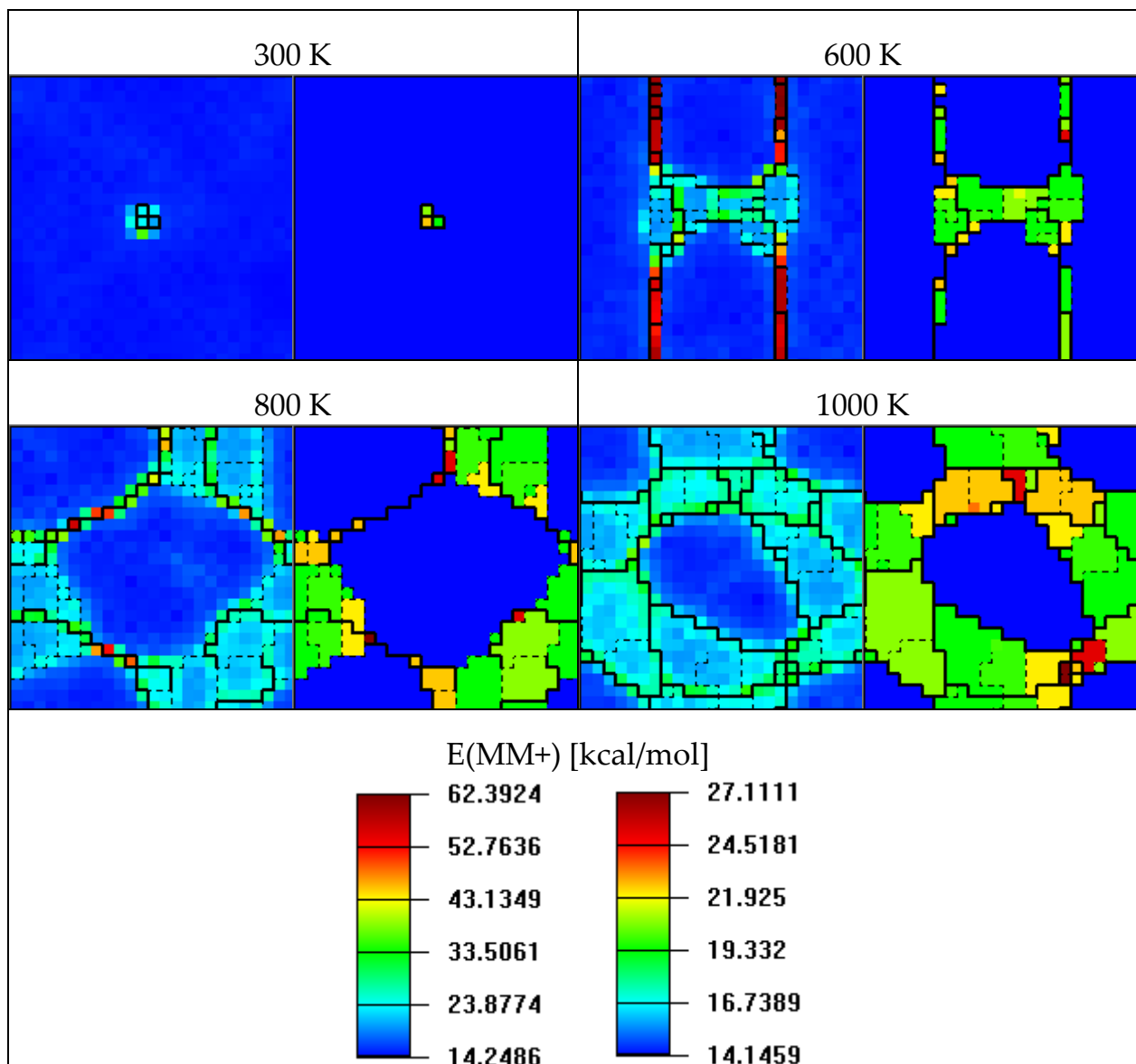
In grafischer Auftragung sind obige Werte in Abbildung 148 zu finden. Während bei 300 K das Molekül den größten Teil der Simulation in der *chair/chair*-Konformation verbringt, steigt der Anteil der *chair/twist*-Konformationen mit der Temperatur. Ein Wechsel der Netztopologie von benachbarten dunkelblauen *chair/chair*-Clustern in getrennte Bereiche findet bei ca. 800 K statt und koinzidiert wieder mit einem Maximum in der RMSG-Kurve. Die RMSDSN-Werte steigen kontinuierlich an, was auf eine kontinuierliche Vergrößerung des Konformationsraums zurückzuführen ist.





**Abbildung 148: Variation der Temperatur – *cis*-Decalin MD, Training der 11 Ringdiederwinkel**

Die obigen Beobachtungen werden bestätigt durch beispielhaft durchgeführte MINIX1- und MINIX2-Rechnungen bei jeder Temperatur. Die Netzkarten sind wiedergegeben in Abbildung 149.



**Abbildung 149:** Beispielnetskarten Gridhyperfläche – Variation der Temperatur – *cis*-Decalin MD, Training der 11 Ringdiederwinkel, links partiell optimiert unter Verwendung der Netzrestraints nach COLORIX1, rechts vollständig optimiert nach COLORIX2

Sämtliche gefundenen Minima auf der Netzkarte sind dabei *chair/chair*-, *twist/chair*- oder *twist/twist*-Kombinationen in den beiden anellierten Ringen zuzuordnen. Eine Zuordnung der Cluster auf der 1000 K Netzkarte wird in Abbildung 150 vorgenommen. Die zur Zuordnung verwendeten Diederwinkel nach Abbildung 147 lauten:

- Ring 1:  $\omega_{1,2} / \omega_{2,3} / \omega_{3,4} / \omega_{4,5} / \omega_{5,6} / \omega_{6,1}$
- Ring 2:  $\omega_{1,7} / \omega_{7,8} / \omega_{8,9} / \omega_{9,10} / \omega_{10,2} / \omega_{2,1}$

Nach dem Schema aus Tabelle 60 werden die Konformationen der beiden Ringe durch ihren Kurznamen bezeichnet. Eine Konformation mit beiden Ringen in *chair*-Konfiguration „1“ heißt so z.B. „11“. Der erste Ring in *chair*-Konformation „1“ und der zweite in *twist* Konformation „f“ hieße „1f“. Wie in

Abbildung 150 zu erkennen ist, sind bei weitem nicht alle Kombinationen ohne Bindungsbruch in der Moleküldynamik erreichbar. Die „12“ oder „21“-Kombinationen existieren im Eingabesatz z.B. nicht.

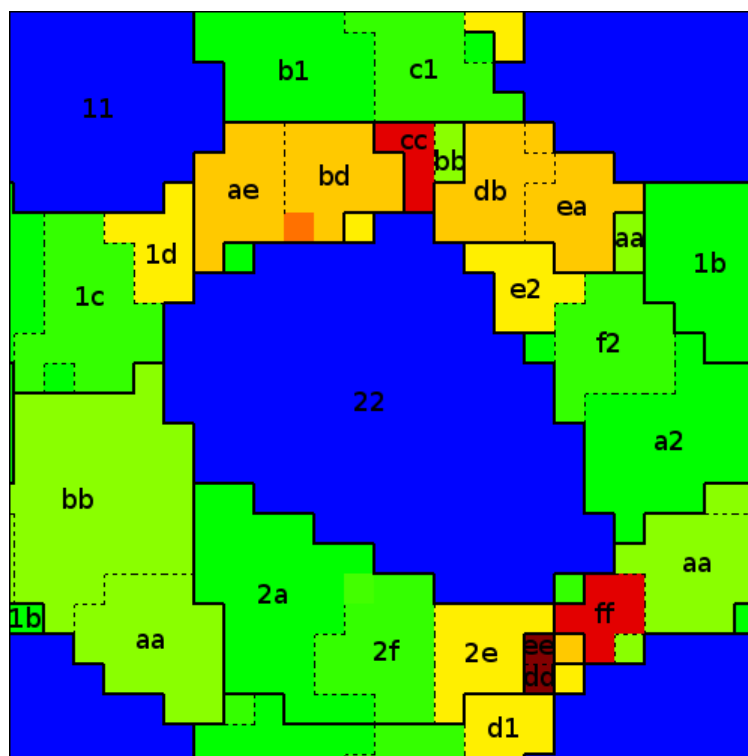


Abbildung 150: *cis*-Decalin - Zuordnung der Cluster auf der 1000 K MD-Kohonenkarte, Training der 11 Diederwinkel nach COLORIX2

### 2.8.7.2 MD *cis*-Decalin, Training 77 zufälliger Distanzen

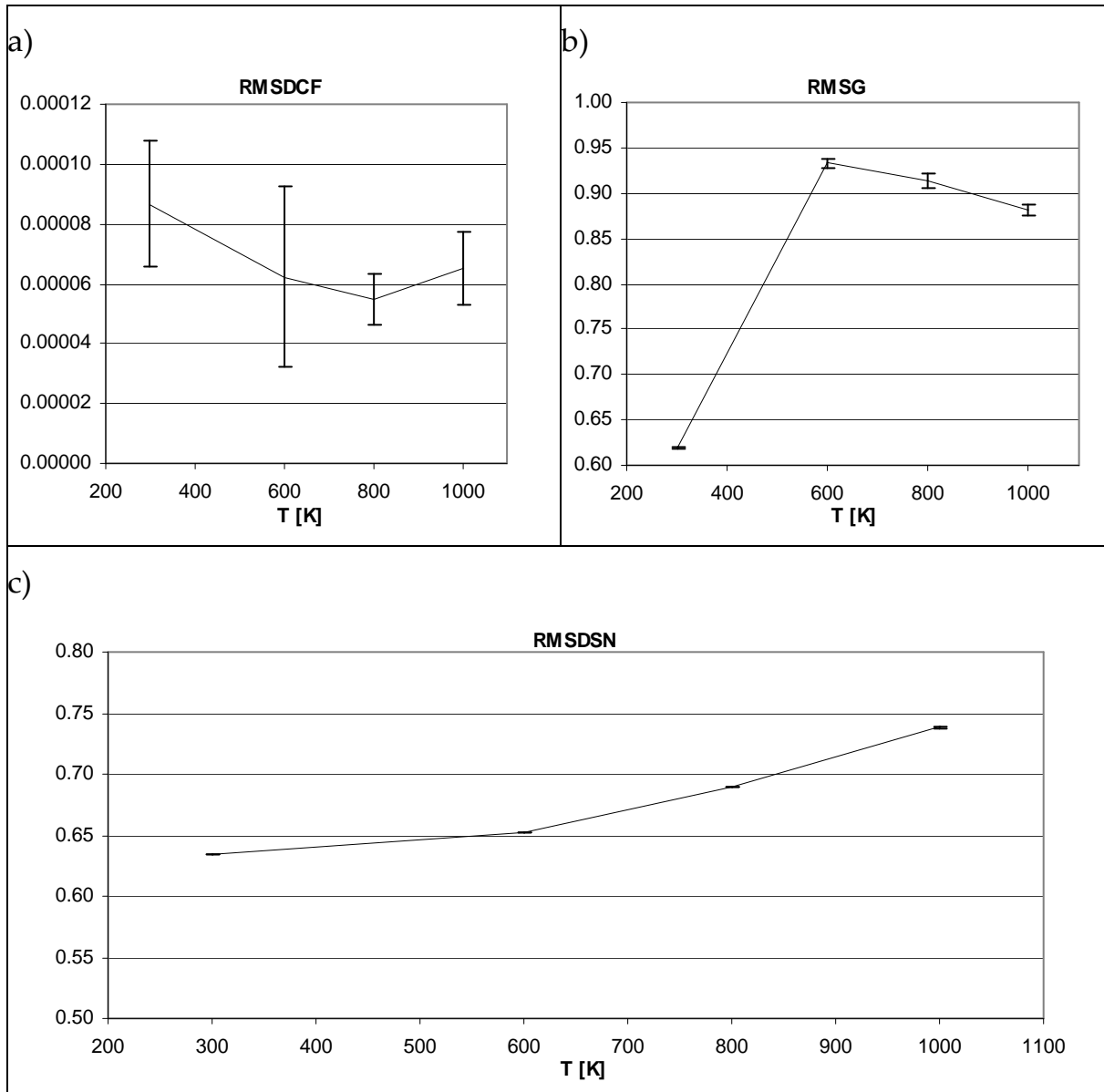
Mit den gleichen Moleküldynamikdaten wie oben werden Kohonennetze mit 77 zufällig ausgewählten CC, CH und HH Distanzen trainiert [25x25 Neuronen; 20 *epochs*;  $\Delta_{\max}(1)=12$ ;  $f_0=0.4$ ; Abstandsfunktion parabolisch, Training der 77 zufällig ausgewählter Distanzen]. Klassifiziert werden die Konformationen zwecks Errechnung der RMSDCF-Werte mit dem Programm CLASSIFY\_SNP\_CONFORMATIONS.EXE.

Tabelle 71: Decalin MD-Simulationen bei unterschiedlichen Temperaturen, Training 77 zufällig ausgewählter Distanzen

T [K]	Ø RMSDCF	$\sigma$	relerr%	min	max	ndata
1000	6.525E-05	1.215E-05	18.62	5.171E-05	8.261E-05	5
800	5.463E-05	8.543E-06	15.64	4.401E-05	6.474E-05	5
600	6.226E-05	3.019E-05	48.50	3.748E-05	1.144E-04	5
300	8.679E-05	2.086E-05	24.03	5.442E-05	1.099E-04	5
T [K]	Ø RMSG	$\sigma$	relerr%	min	max	ndata
1000	0.88169	0.00549	0.62	0.87206	0.88542	5
800	0.91281	0.00784	0.86	0.90462	0.92278	5
600	0.93319	0.00461	0.49	0.92683	0.93944	5
300	0.61960	0.00060	0.10	0.61890	0.62028	5

T [K]	Ø RMSDSN	$\sigma$	relerr%	min	max	ndata
1000	0.73812	0.00011	0.02	0.73793	0.73821	5
800	0.68936	0.00017	0.03	0.68914	0.68958	5
600	0.65227	0.00018	0.03	0.65210	0.65256	5
300	0.63475	0.00007	0.01	0.63465	0.63480	5

Grafisch aufgetragen sind diese Werte in Abbildung 151



**Abbildung 151: Variation der Temperatur – *cis*-Decalin MD, Training 77 zufällig ausgewählter Distanzen**

RMSG-Werte und RMSDSN-Werte zeigen den gleichen Trend, und auch die Netzkarten in Abbildung 152 zeigen ähnliche Netztopologien wie beim Training der Diederwinkel. Von allen drei Trainingsätzen machen die Netzkarten den klarsten Eindruck.

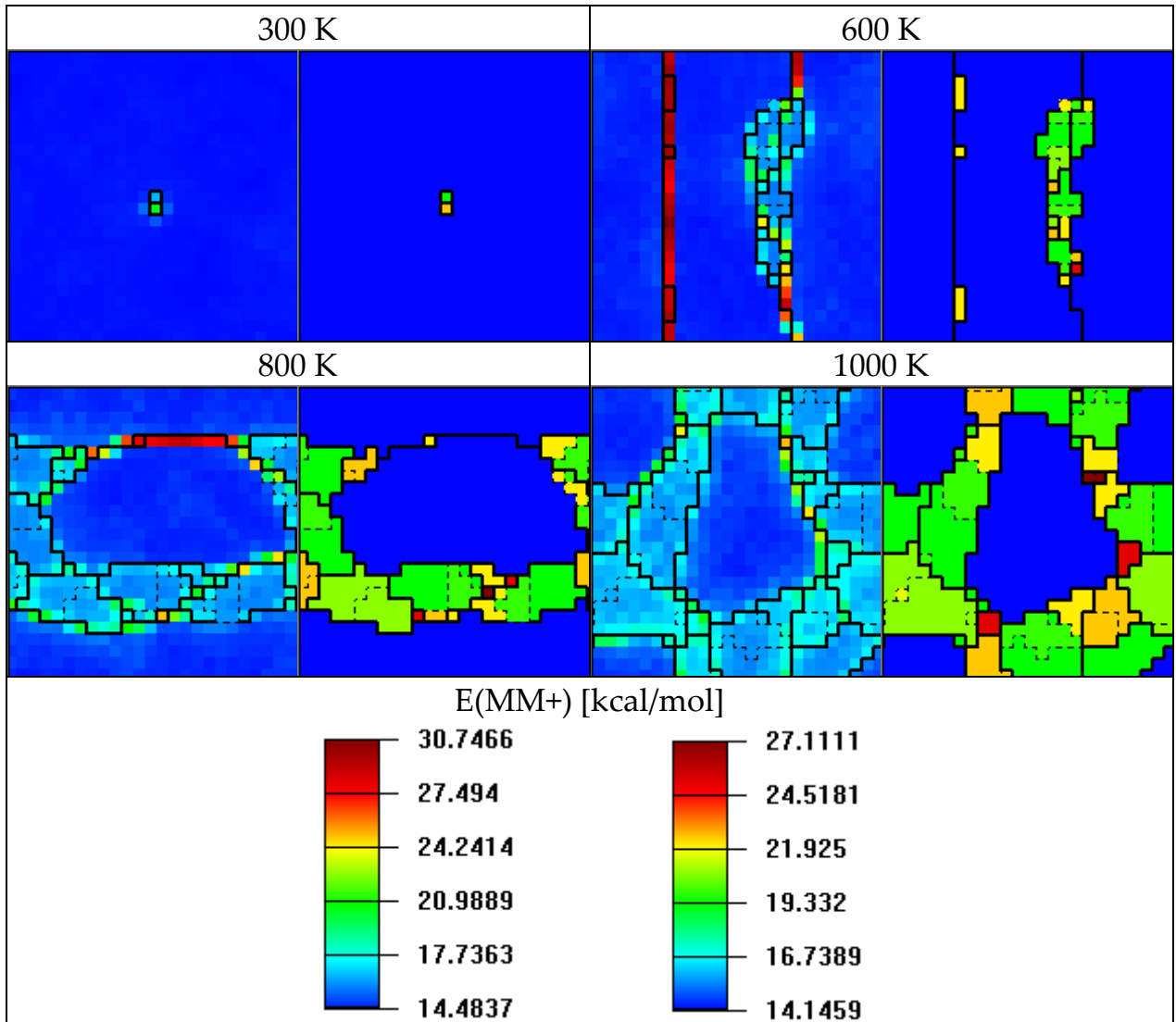


Abbildung 152: Beispielnetzarten Gridhyperfläche – Variation der Temperatur – *cis*-Decalin MD, Training 77 zufällig ausgewählter Distanzen, links partiell optimiert unter Verwendung der Netzrestraints nach COLORIX1, rechts vollständig optimiert nach COLORIX2

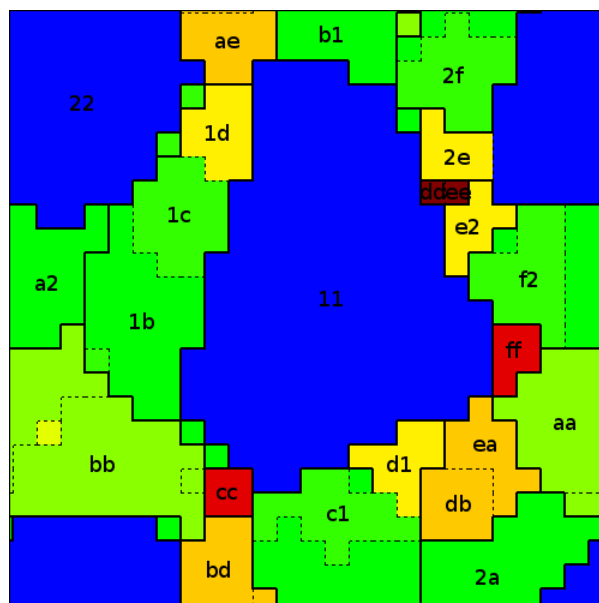


Abbildung 153: *cis*-Decalin - Zuordnung der Cluster auf der 1000 K MD-Kohonenkarte, Training 77 zufällig ausgewählter Distanzen

In Abbildung 153 werden nach gleichem Schema wie im vorherigen Kapitel Zuordnungen der Cluster auf der 1000 K MD-Netz Karte vorgenommen. Die Netz Karte scheint leicht unfragmentierter als beim Training der Diederwinkel.

### 2.8.7.3 MD cis-Decalin, Training der Schweratomdistanzen

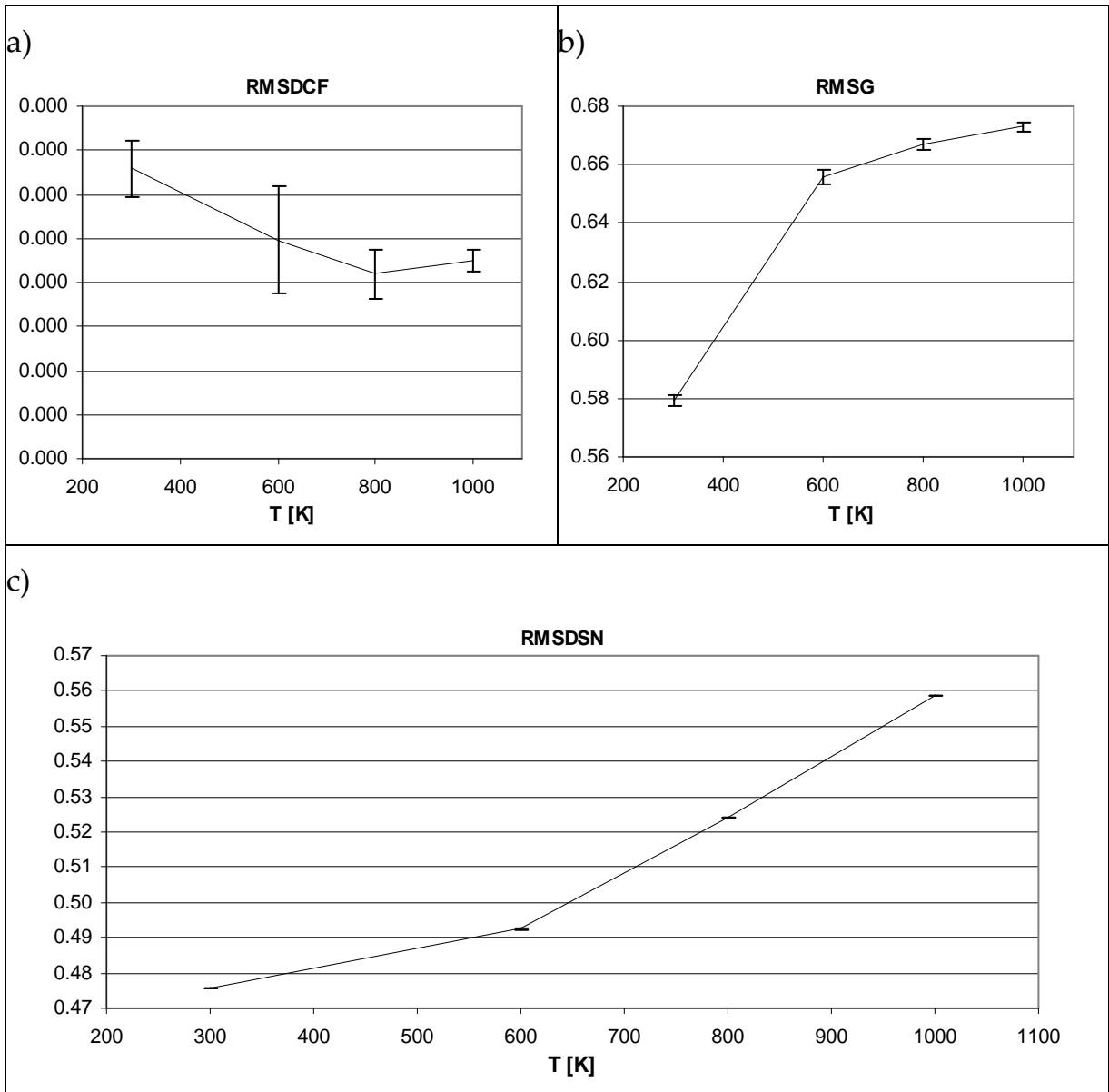
Mit den gleichen Moleküldynamikdaten wie oben werden Kohonennetze mit sämtlichen CC-Abständen trainiert. [25x25 Neuronen; 20 *epochs*;  $\Delta_{\max}(1)=12$ ;  $f_0=0.4$ ; Abstandsfunktion parabolisch, Training der 45 CC-Distanzen]. Die RMSDCF-Werte können an dieser Stelle sinnvoll angegeben werden, da durch die Schweratomdistanzmatrix nicht unterscheidbare Konformationen in den zugrunde liegenden Moleküldynamiken ohne Bindungsbruch nicht auftreten können.

**Tabelle 72:** Decalin MD-Simulationen bei unterschiedlichen Temperaturen, Training der CC-Distanzen

T [K]	Ø RMSDCF	$\sigma$	relerr%	min	max	ndata
1000	9.000E-05	4.816E-06	5.35	8.522E-05	9.608E-05	5
800	8.385E-05	1.133E-05	13.52	6.880E-05	1.004E-04	5
600	9.917E-05	2.417E-05	24.37	7.274E-05	1.247E-04	5
300	1.318E-04	1.284E-05	9.74	1.169E-04	1.516E-04	5
T [K]	Ø RMSG	$\sigma$	relerr%	min	max	ndata
1000	0.67293	0.00144	0.21	0.67077	0.67471	5
800	0.66674	0.00189	0.28	0.66364	0.66877	5
600	0.65551	0.00252	0.38	0.65130	0.65773	5
300	0.57928	0.00206	0.35	0.57693	0.58191	5
T [K]	Ø RMSDSN	$\sigma$	relerr%	min	max	ndata
1000	0.55856	0.00005	0.01	0.55852	0.55865	5
800	0.52402	0.00010	0.02	0.52387	0.52412	5
600	0.49249	0.00009	0.02	0.49239	0.49262	5
300	0.47570	0.00014	0.03	0.47552	0.47586	5

Die grafische Auftragung zeigt steigende RMSG- und RMSDSN-Werte in Abbildung 154. Im Vergleich zu den obigen beiden Trainingsreihen gibt es kein Maximum im RMSG-Graphen und auch die Netztopologie macht keine großen Veränderungen im Verlaufe der Temperaturerhöhung durch.





**Abbildung 154: Variation der Temperatur – *cis*-Decalin MD, Training der CC-Distanzen**

Die erhaltenen Beispielnetskarten, siehe Abbildung 155, sind unruhiger und „zerfranster“ als die Netskarten aus den vorherigen Kapiteln. Trotzdem ist die Auswertung der Hyperfläche am übersichtlichsten.

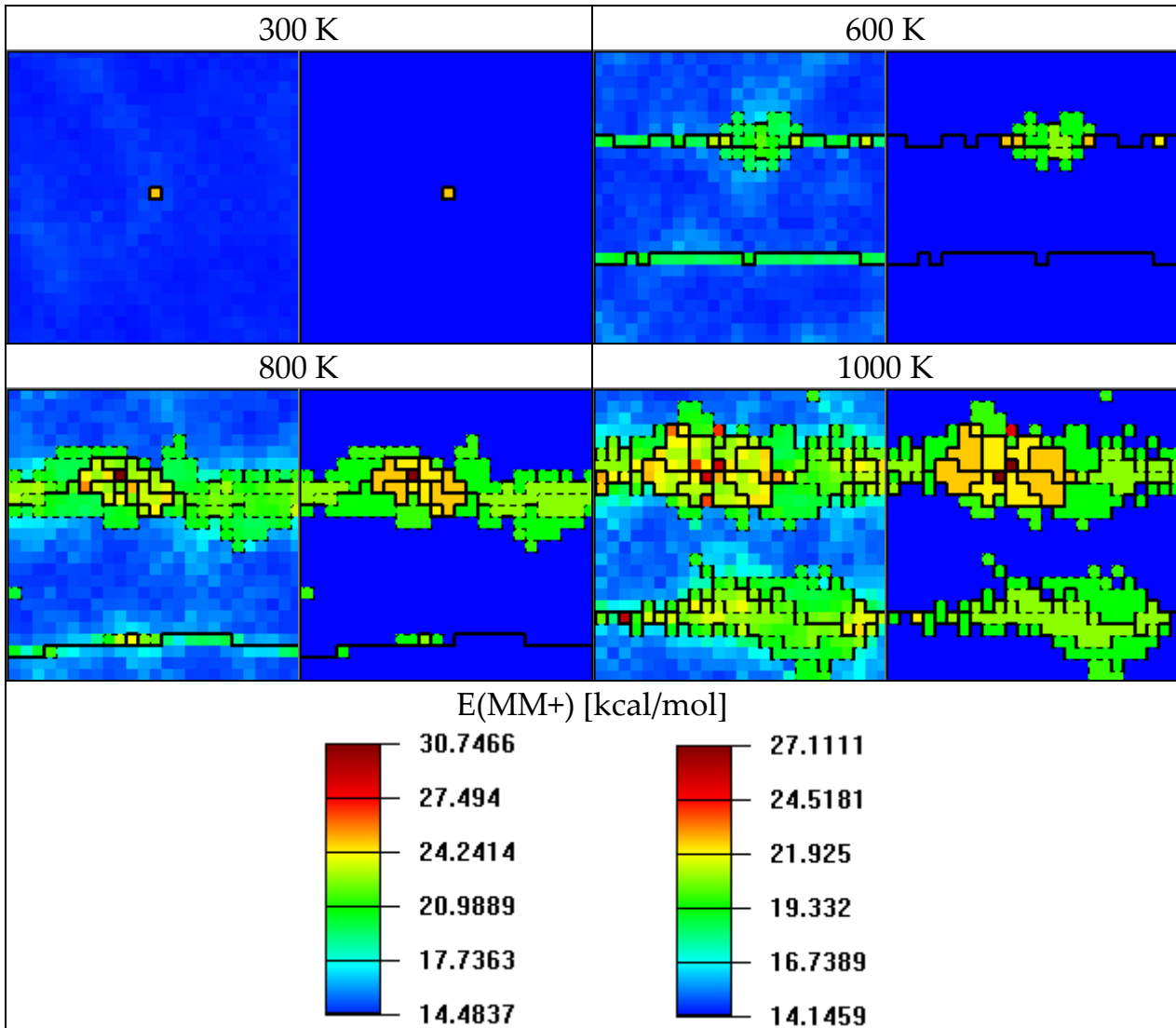


Abbildung 155: Beispielnetzarten – Variation der Temperatur – *cis*-Decalin MD, Training der CC-Distanzen, links COLORIX1, rechts COLORIX2

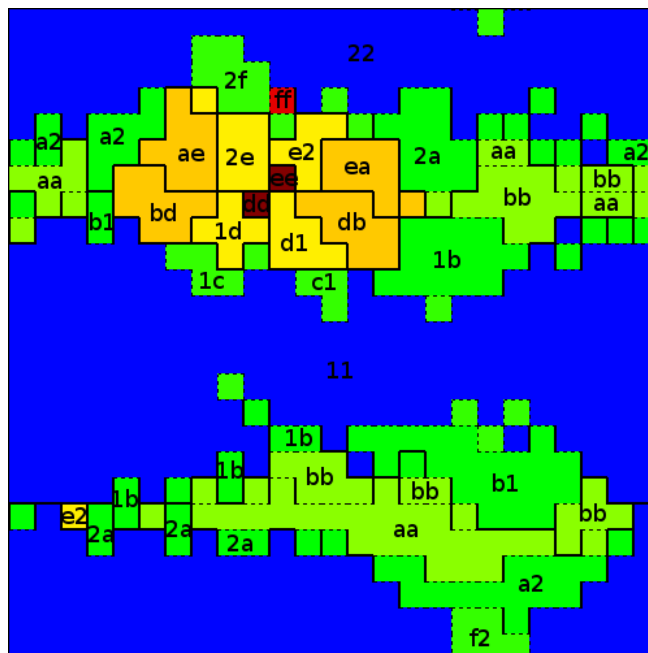


Abbildung 156: *cis*-Decalin - Zuordnung der Cluster auf der 1000 K MD-Kohonenkarte, Training der CC-Distanzen

In Abbildung 156 werden nach gleichem Schema wie in den vorherigen Kapiteln Zuordnungen der Cluster auf der 1000K MD-Netz Karte vorgenommen. Die Pseudosymmetrie der Hyperfläche kommt in dieser Darstellung am besten zum Ausdruck, allerdings ist die Netz Karte stark fragmentiert mit einzelnen Ausreißern in den ansonsten homogenen Flächen.

### 2.8.8 Dibromcycloocten

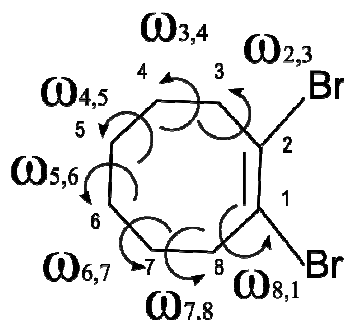


Abbildung 157: CC-Diederwinkel im Dibromcycloocten

Das Dibromcyclooctenmolekül dient als erster Anwendungsfall, um Hyperflächen, welche auf verschiedene Weisen erhalten wurden, miteinander zu vergleichen. Dazu wird eine relativ kurze MD-Rechnung einer Konformationsuche mit dem Programm HYPERCHEM gegenübergestellt.

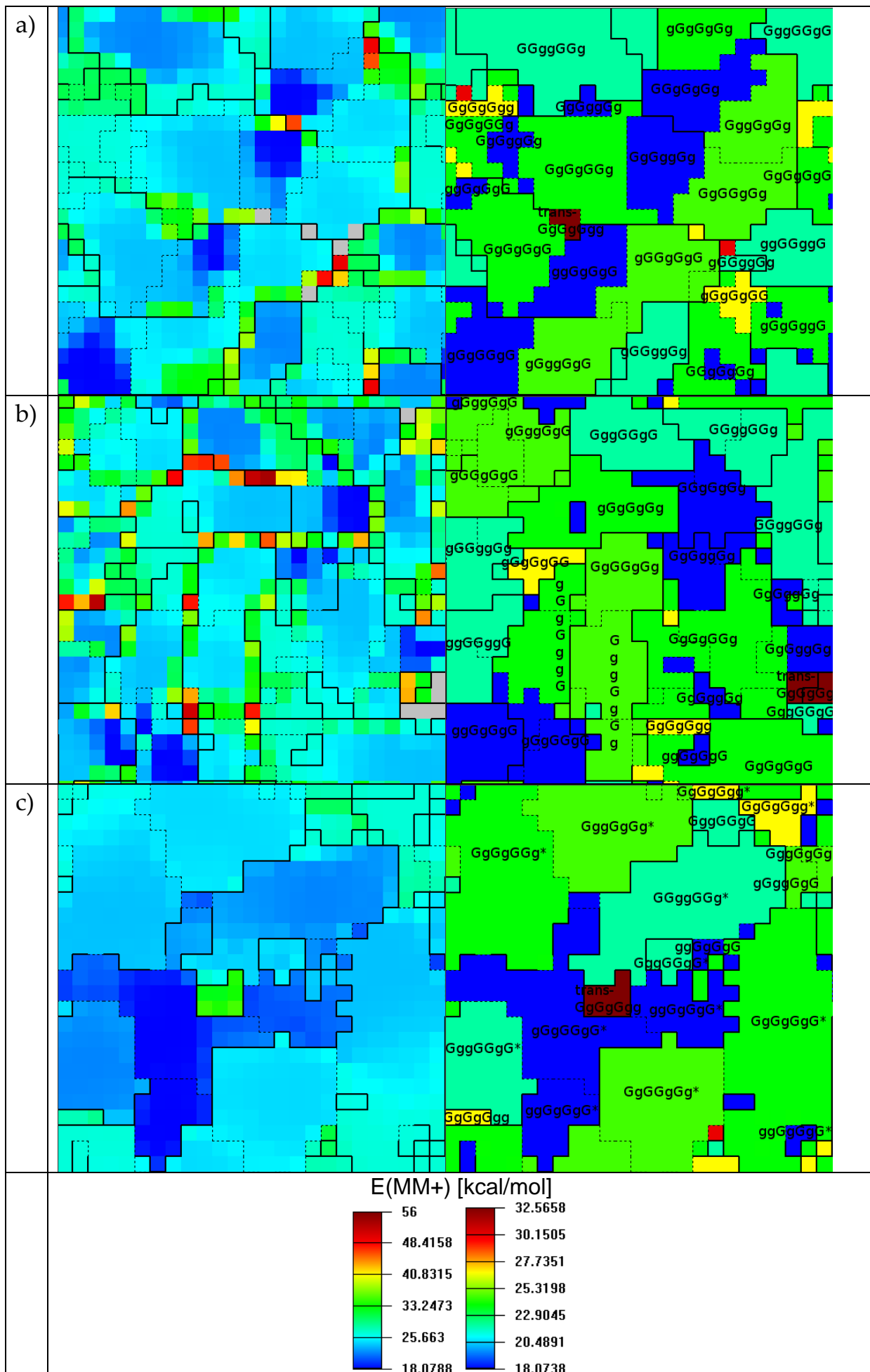
#### 2.8.8.1 Konformationssuche (HYPERCHEM conformational search)

Die Daten im Konformerensatz der systematischen Suche werden erzeugt mit dem Konformationssuchmodul (*Compute/Conformational Search*) des Programms HYPERCHEM<sup>[43],[117]</sup> und mit Hilfe des Programms HCS2SNP (siehe Kapitel 5.3.10) in ein für die NEURO-SUITE lesbares Format umgewandelt.

Die Suche wird unter gleichzeitiger Variation der Diederwinkel ( $\omega_{3,4}$ ,  $\omega_{4,5}$ ,  $\omega_{5,6}$ ,  $\omega_{6,7}$ ,  $\omega_{7,8}$ ) durchgeführt. Die 284 erhaltenen Strukturen sind optimiert mit dem Kraftfeld MM+, genaue Parameter im *supplementary material*.<sup>[118]</sup>

Die Strukturen aus obiger Konformationssuche werden trainiert unter Verwendung 3 verschiedener Sätze an Eingabevektoren, siehe Abbildung 158.

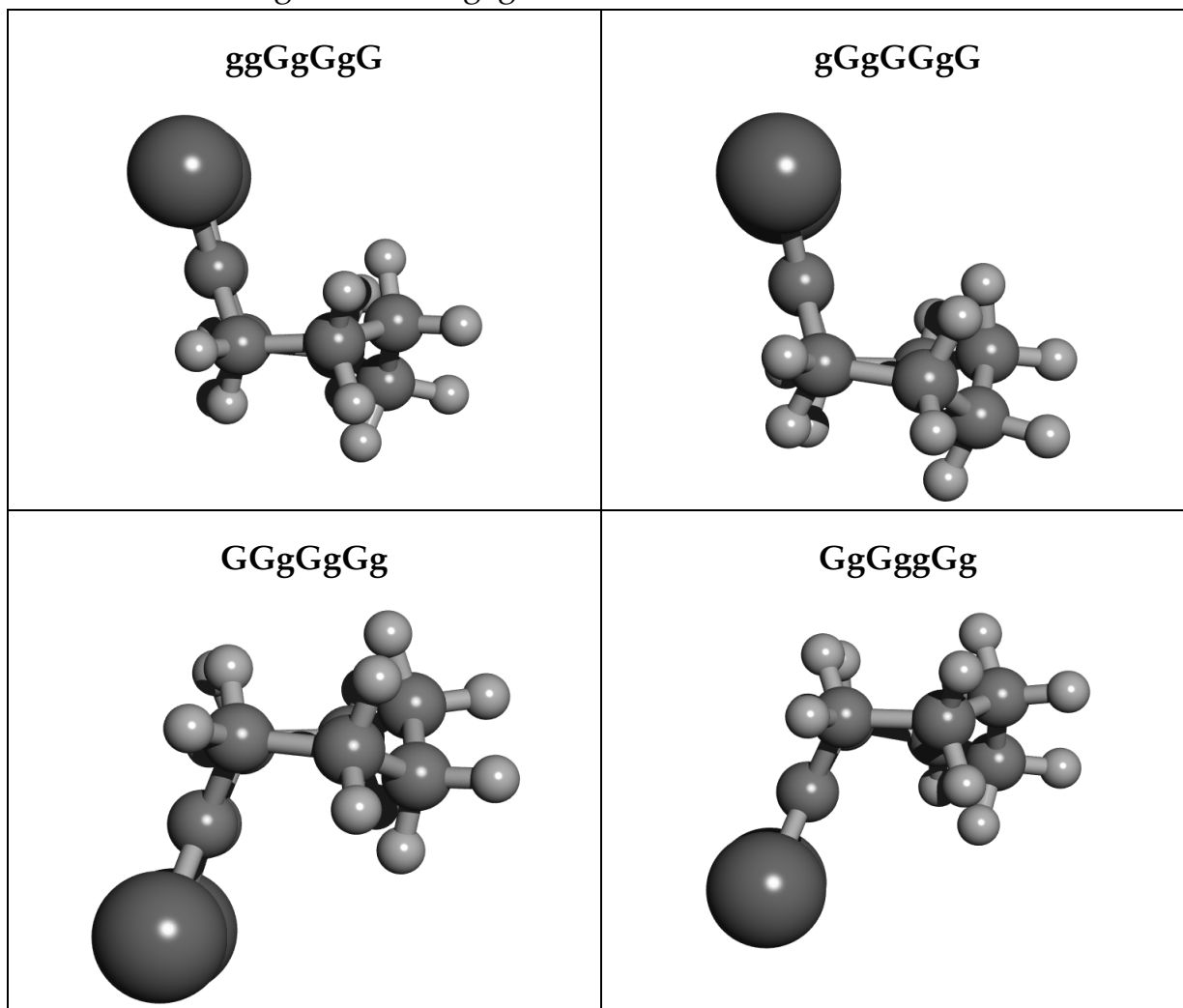
[25x25 Neuronen; 10000 *epochs*;  $\Delta_{\max}(1)=6$ ;  $f_0=0.1$ ; parabolisch; Abstandsfunktion parabolisch]



**Abbildung 158:** Netzkarten Konformationssuche Dibromcycloocten, links partiell optimiert unter Verwendung der Netzrestraints nach COLORIX1, rechts vollständig optimiert nach COLORIX2  
 a) Training der Diederwinkel aus Abbildung 157  
 b) Training 51 zufällig ausgewählter Atomabstände  
 c) Training der 28 Ringatomabstände

In Abbildung 158 c) kommt es wieder zur Äquivalenz von Enantiomeren auf der Netzoberfläche. So entspricht z.B. die GggGgGg Konformation von den trainierten Abständen her seinem gGGgGgG Enantiomer. Derartige Bereiche sind in Abbildung 158 c) mit einem Stern gekennzeichnet. Aus dem Nichtvorhandensein eines solchen Sterns kann jedoch nicht geschlossen werden, dass das entsprechende Enantiomer im Trainingsatz nicht vorkommt, da die Zuordnung der Enantiomere zu den Neuronen stochastisch erfolgt.

Die Molekülstrukturen zu den vier dunkelblauen globalen Minimumsclustern sind in Abbildung 159 wiedergegeben.



**Abbildung 159:** globale Minimumsstruktur aus Abbildung 158 a), „Numerationenantiomerie“ (oben ↔ unten) und „Numerationsdiastereomerie“ (links ↔ rechts)

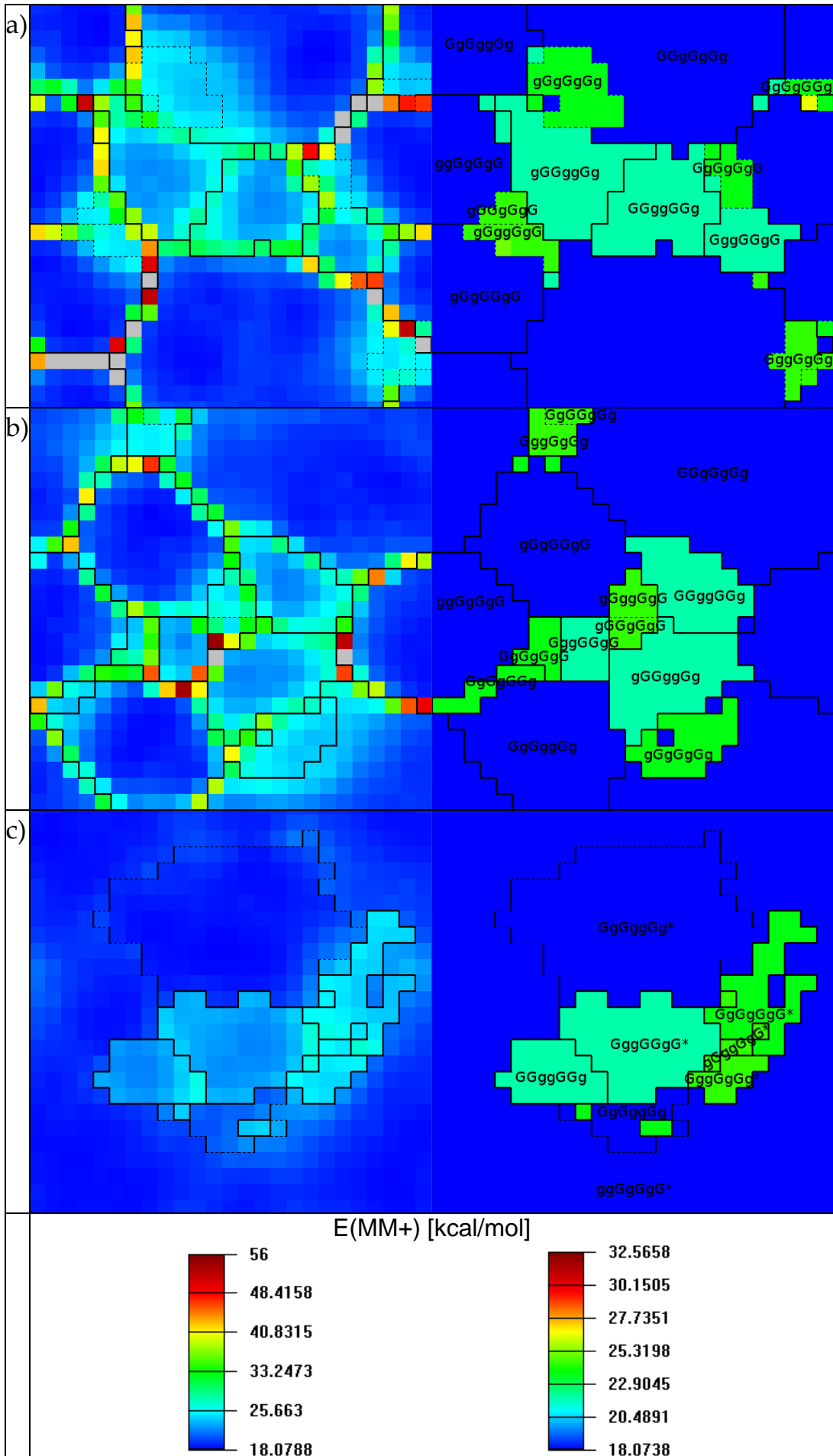
Die höchste Energie auf der Netzkarte erreicht mit 32.56 kcal/mol, bzw. einer relativen Energie in Bezug auf das globale Minimum von 14.49 kcal/mol eine Struktur mit *trans*-konfigurierter Doppelbindung, die aufgrund der variierten Diederwinkel eigentlich nicht erwartet wird.

### **2.8.8.2 MD 800 K**

Als weitere Möglichkeit der Generierung von Hyperflächen wird der systematischen Konformationsuche aus obigem Kapitel eine recht kurze Moleküldynamik gegenübergestellt. Um den abgesuchten Konformationsraum drastisch zu vergrößern, wird die Moleküldynamik bei 800 K durchgeführt, auch wenn die infolge der hohen potentiellen Energien erhaltenen Auslenkungen aus der Gleichgewichtslage für Abstände und Winkel im Molekül deutlich außerhalb des zulässigen Bereichs von Kraftfeldern im allgemeinen liegen. Die Verwendbarkeit dieser Daten zum Training eines Kohonnennetzes wird in den vorhergehenden Kapiteln durch Erhalt realistischer Hyperflächen gerechtfertigt.

Die erhaltenen Netzkarten unter Verwendung der gleichen Trainingsvektoren wie im vorhergehenden Kapitel sind in Abbildung 160 wiedergegeben.

(Kraftfeld HYPERCHEM MM+, in vacuo, 20000 Strukturen, Simulationsdauer 200 ps, Aufruf von FIX\_ROTATION.EXE bei je 10 ps, *step size* 0.001 ps, *datafreq=1/10 data steps<sup>-1</sup>*)



**Abbildung 160:** Netzkarten MD 800 K Dibromcycloocten, links partiell optimiert unter Verwendung der Netzrestraints nach COLORIX1, rechts vollständig optimiert nach COLORIX2  
a) Training der Diederwinkel aus Abbildung 157  
b) Training 51 zufällig ausgewählter Atomabstände  
c) Training der 28 Ringatomabstände

Wie insbesondere in den Abbildung 158 b) und c) zu sehen ist, ist der abgedeckte Konformationsraum deutlich kleiner als bei der systematischen Suche, insbesondere sind keine Konformationen mit einer Energie über 24.61 kcal/mol (6.54 kcal/mol relativ zum globalen Minimum) im Trainingsatz enthalten.

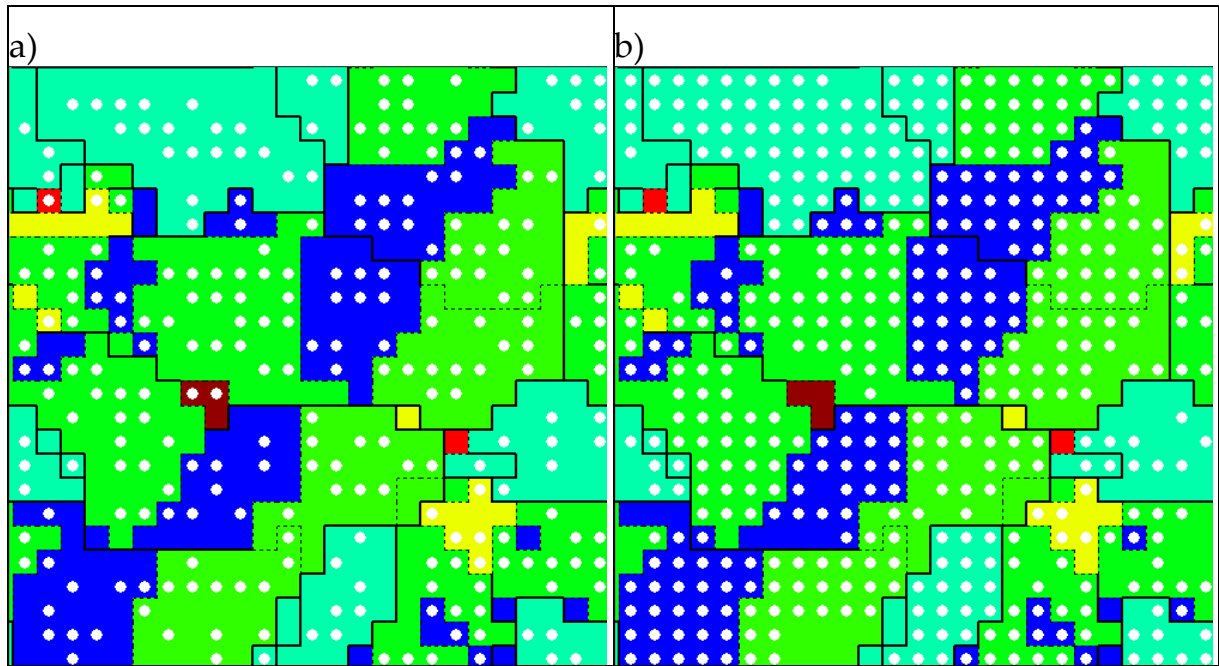
### 2.8.8.3 Konformationsräume im Vergleich

Eine interessante Fragestellung bei jeder Methode zur Generierung von Strukturensambles betrifft den Umfang der Abbildung des Konformationsraums. Generell kann keine Methode für eine vollständige Abdeckung für beliebige Moleküle garantieren. So hat Dill eine Gleichung zur Abschätzung der oberen Grenze der thermisch zugänglichen Konformationen in einem Protein entwickelt<sup>[119]</sup>. Unter Verwendung dieser Gleichung erhält man für ein Protein mit 150 Aminosäuren eine Obergrenze von  $4 \cdot 10^{34}$  Konformationen, was deutlich macht, dass eine vollständige Betrachtung weder möglich noch wünschenswert ist.

Klassische Verfahren der Analyse von Konformationsräumen haben mit dem Vergleich zweier hochredunder Ensembles, wie sie z.B. aus MD-Rechnungen erhalten werden, Probleme, wenn die Konformere nicht relativ gleichförmig über die gesamte Hyperfläche verteilt sind. Es wird an dieser Stelle eine einfache und effiziente Methode zur Verwendung der Kohonnenkarten vorgestellt, welche diese Probleme nicht hat.

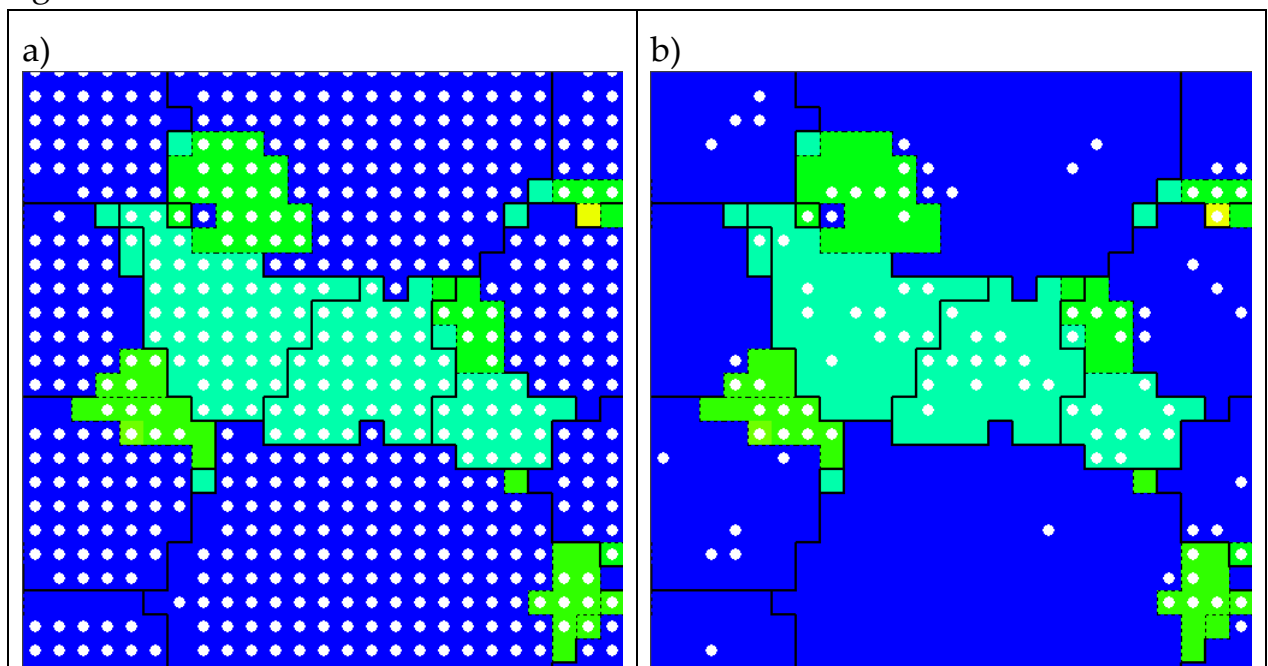
Dazu wird der Hypersurface Explorer, beschrieben in Kapitel 2.8.1.4, eingesetzt. Auf die Netzkarte der Konformationssuche über die Torsionswinkel, siehe Abbildung 158 a) werden in Abbildung 161 die Konformationsensemble aus der Konformationssuche (*HYPERCHEM conformational search*) sowie der MD-Simulation projiziert und jedes Konformer als weißer Punkt dargestellt. Die Moleküldynamikrechnung erzielt dabei Treffer auf praktisch allen Clustern der Netzkarte mit Ausnahme der hochenergetischen *trans*-Konformation.





**Abbildung 161:** Projektion (weiße Punkte) auf die Netzkarte der Konformationssuche (HYPERCHEM *conformational search*), Training der Diederwinkel  
 a) Strukturen aus der Konformationssuche (*conformational search*)  
 b) Strukturen aus der MD-Rechnung

Es kann konstatiert werden, dass der Konformationsraum der Konformationssuche weitgehend durch die MD-Rechnung abgedeckt ist. In Abbildung 162 werden zum Vergleich die beiden Trainingsensembles auf die Netzkarte der MD-Rechnung (wiederum die Diederkarte) projiziert. Sämtliche größeren Cluster auf der Karte werden durch beide Datensätze ziemlich „mittig“ getroffen.



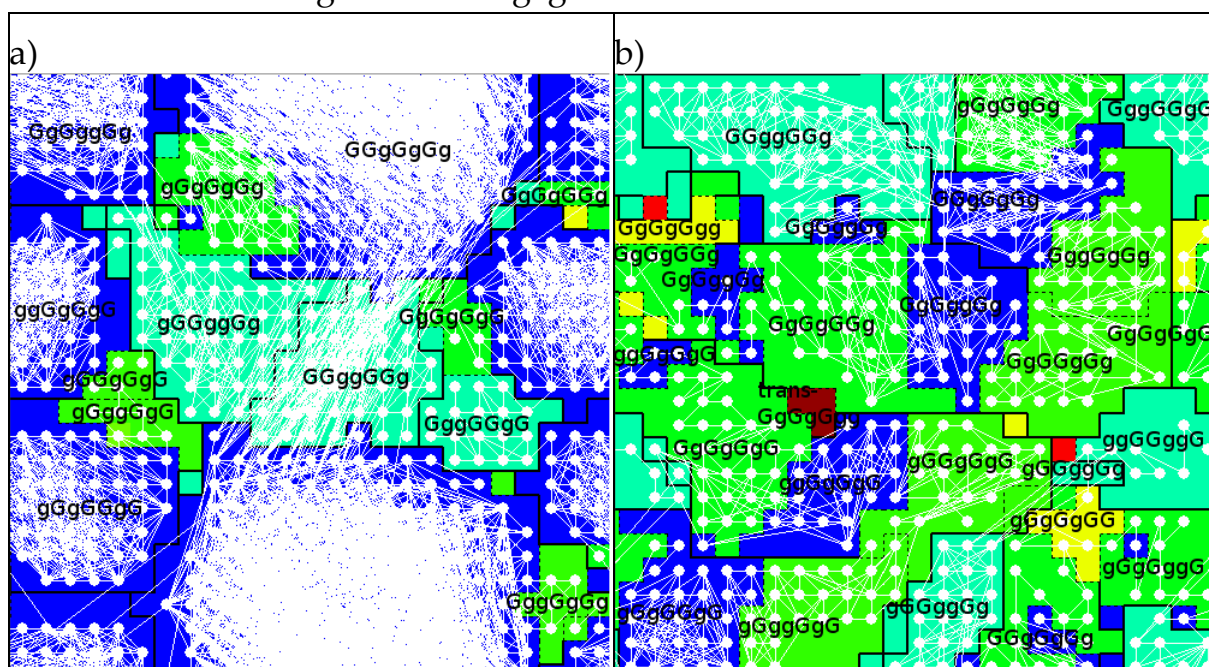
**Abbildung 162:** Projektion (weiße Punkte) auf die Netzkarte der MD-Rechnung, erhalten durch Training der Diederwinkel  
 a) Strukturen aus der MD-Rechnung  
 b) Strukturen aus der Konformationssuche (*conformational search*)

Dies lässt den Schluss zu, dass der Konformationsraum der MD-Rechnung komplett in der Konformationsuche enthalten ist.

#### 2.8.8.4 Topologiebeziehungen

Nachbarschaftsbeziehungen zwischen den Konformationen lassen sich mit dem *Hypersurface Explorer* darstellen, indem eine ausreichend lange MD-Rechnung auf die Hyperfläche wie oben projiziert wird, wobei die Punkte konsekutiver Strukturen mit einer Geraden verbunden werden. Im ursprünglichen dreidimensionalen Koordinatenraum benachbarte Strukturen lassen sich dann durch Übergänge, präsentiert durch Verbindungslinien auf der Netzkarte, erkennen.

Die Projektion der MD-Rechnung bei 800 K inklusive dieser Verbindungslinien ist in Abbildung 163 wiedergegeben.



**Abbildung 163:** Projektion der MD-Trajektorie auf die Netzkarten, inklusive Verbindungslinien

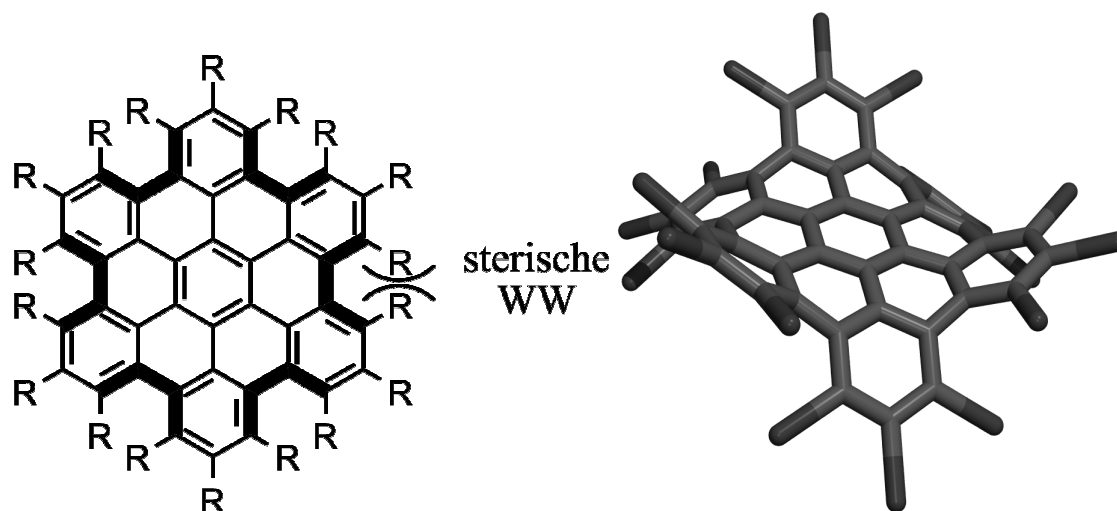
a) Netzkarte der MD-Rechnung (originaler Datensatz)

b) Netzkarte der Konformationsuche

Als Ergebnis kann festgehalten werden, dass Umwandlungen zwischen den vier tiefblauen globalen Minimumskonformeren, selbst bei unmittelbarer Nachbarschaft, in der Regel über grüne Konformationen geschieht, was auf hohe energetische Barrieren schließen lässt. Ein Artefakt der unvollständigen Kalibrierung der MD-Rechnung sind die langen Linien über die GGggGGg-Konformation in Abbildung 163a). Diese sind keine Konformationsübergänge sondern Verbindungen von „Randbereichen“ der GGgGgGg-Konformation mit sich selbst, welche aufgrund des übergroßen Clusters auf kürzester Linie quer über die Karte erfolgt.

### 2.8.9 Hexabenzocoronene (HBC)

Eine der ersten Anwendungen der Neuroprogrammiersuite ist die Konformationsanalyse der Hexabenzocoronene in Kooperation mit F. Dötz von der Universität Mainz, der im Rahmen seiner Dissertation<sup>[120]</sup> die ersten vollständig substituierten Hexabenzocoronene synthetisierte.



**Abbildung 164:** persubstituiertes HBC: links Schema, rechts 3D-Darstellung der globalen Minimumsstruktur(+-+-+):  
 a) R=Br: Modellsubstanz zur strukturellen Analyse der Grundgerüstkonformere.  
 b) R=OMe: synthetisiertes Derivat, Strukturbestimmung durch Röntgenstrukturanalyse in der Dissertation von F. Dötz.<sup>[120]</sup>

Ziel der zitierten Arbeit ist die Induktion von dreidimensionaler Verformung in der normalerweise planaren Graphitstruktur durch räumliche Überfüllung. Die eingezeichneten sterischen Wechselwirkungen der Reste **R** in Abbildung 164 bewirken eine Verdrillung aus der Planarität, welche sich in einer Abweichung aus der planaren 0°-Konfiguration der fett eingezeichneten Diederwinkel manifestiert, im Folgenden *bay*-Diederwinkel benannt.

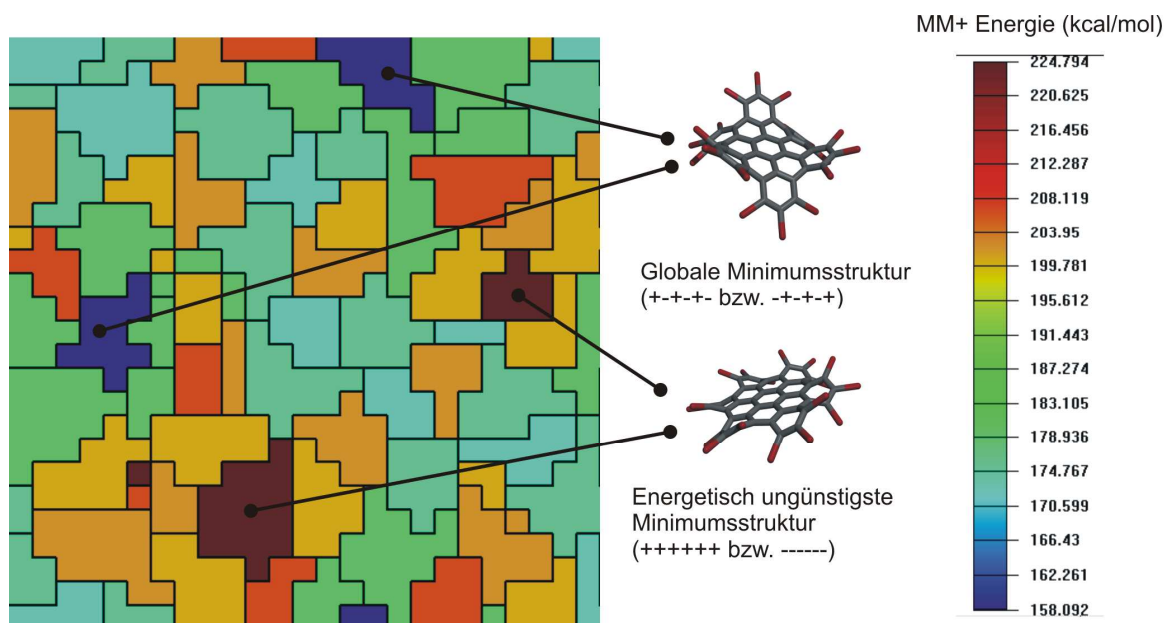
Die Konformationsanalyse des synthetisierten permethoxy-HBC wird erschwert durch die Rotation der Methylether-Reste. Für eine systematische Untersuchung der Gerüstkonformere wird zunächst ein hypothetisches perbromiertes HBC untersucht.

#### 2.8.9.1 Perbromo-HBC

Bromatome als Reste **R** haben einen deutlich größeren Raumbedarf als Methoxygruppen, außerdem gibt es keine strukturellen Auswirkungen durch Drehung der Reste. Dies führt zu einer erheblich einfacheren Hyperfläche, auf welcher die Konformationen durch Angabe der Vorzeichen der *bay*-Winkel eindeutig definiert sind, siehe rechte Seite der Abbildung 164. Unter Verwen-

dung des Skriptes SYSTEMATIC\_SEARCH\_PERBROMO\_HBC.TK werden sämtliche 64 Konformationen generiert und damit eine Kohonenkarte trainiert.

[25x25 Neuronen; 2000 *epochs*;  $\Delta_{\max}(1)=12$ ;  $f_0=0.15$ ; Abstandsfunktion parabolisch, Training *bay*-Diederwinkel]



**Abbildung 165:** Kohonenkarte des Perbromo-HBCs<sup>[120]</sup>

Die komplett alternierende Struktur +--+ zeigt die niedrigste Energie und gleichzeitig die größte Auslenkung aus der Planarität.

### 2.8.9.2 Permethoxy-HBC

Das Permethoxy-HBC konnte von F. Dötz synthetisiert und eindeutig per Röntgenstrukturanalyse charakterisiert werden und ist in Abbildung 166 wiedergegeben. Die Hyperfläche ist aufgrund der möglichen Isomerie durch Rotation der Methylethergruppen recht kompliziert. Als Trainingsparameter werden wie im Falle des Perbromo-Derivats die *bay*-Diederwinkel verwendet, auch wenn die Hyperfläche damit nicht vollständig beschrieben wird. Als Rechtfertigung kann angeführt werden, dass die Barrieren für Konformationsänderungen in der *bay*-Region erfahrungsgemäß deutlich höher liegen als die Rotationsbarrieren bei den frei rotierbaren Einfachbindungen der Methoxygruppen. Damit stehen die Cluster der Gerüstisomere auf der Hyperfläche für eine Gruppe von Rotationsisomeren.

Zur Erzeugung der Hyperfläche werden zwei Methoden verwendet, zum einen das Konformationssuchmodul des Programms HYPERCHEM, wobei im Rahmen von 1000 Optimierungen 106 Strukturen gefunden wurden<sup>[121]</sup>, zum anderen eine MD-Simulation.

### 2.8.9.2.1 Konformationsanalyse (HYPERCHEM conformational search)

Die Trainingsdaten der HYPERCHEM *conformational search* werden in der Kohonenkarte aus Abbildung 166 dargestellt. [25x25 Neuronen; 1000 *epochs*;  $\Delta_{\max}(1)=12$ ;  $f_0=0.15$ ; Abstandsfunktion parabolisch, Training *bay-Diederwinkel*]

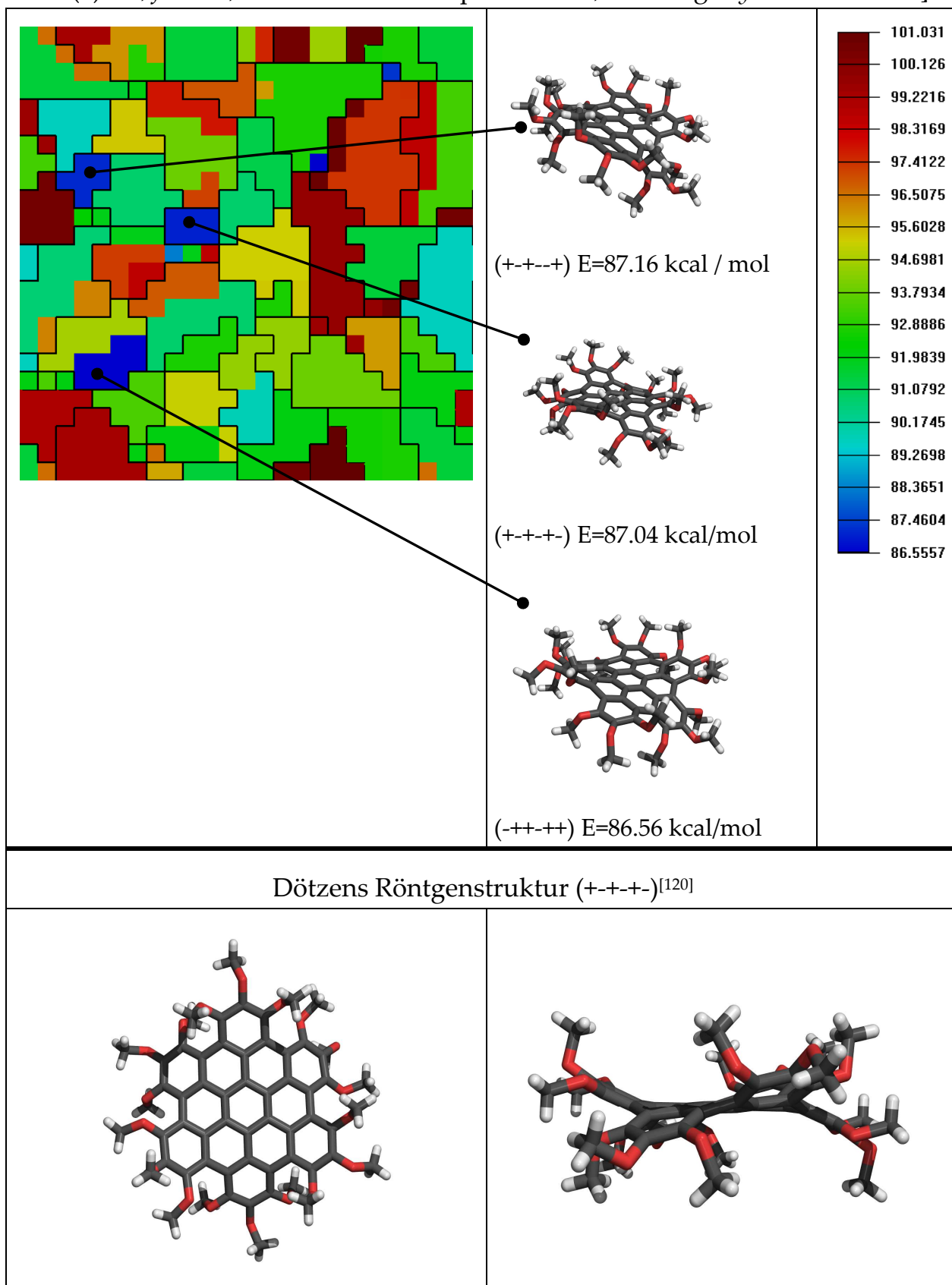


Abbildung 166: Kohonenkarte des Permethoxy-HBCs mit Zuordnung und Röntgenstruktur<sup>[120]</sup>

Eingezeichnet in obiger Darstellung sind die drei energetisch niedrigsten Minima der Analyse. Die Analyse ist mit lediglich eintausend durchgeführten Optimierungen als Einzelschritte der Konformationsanalyse weit davon entfernt, vollständig zu sein, was auch aus dem Fehlen der zugehörigen Spiegelisomere geschlossen werden kann. Im Vergleich zum Perbromo-HBC aus Abbildung 165 fallen die dort ausgeprägten Energieunterschiede zwischen den Gerüstisomeren wesentlich niedriger aus. Dies kann auf den niedrigeren Raumbedarf und die höhere Flexibilität der Methoxygruppen zurückgeführt werden. Interessanterweise entspricht das mittlere (+--+)-Minimum aus Abbildung 166 weitgehend der tatsächlich gefundenen Röntgenstruktur. Kristallpackungseffekte bestimmen wohl die Lage der Methoxygruppen.

### 2.8.9.2.2 Moleküldynamik

Die Laufzeitparameter der MD-Rechnung betragen: [Kraftfeld HYPERCHEM MM+, in vacuo, 1100 K, 20000 Strukturen, Simulationsdauer 1000 ps, Aufruf von FIX\_ROTATION.EXE bei je 50 ps, *step size* 0.001 ps, *datafreq*=1/50 *data steps*<sup>-1</sup>]. Aus den Trainingsdaten der MD-Rechnung wird die Kohonenkarte aus Abbildung 167 erhalten. [25x25 Neuronen; 100 *epochs*;  $\Delta_{\max}(1)=12$ ;  $f_0=0.15$ ; Abstandsfunktion parabolisch, Training *bay*-Diederwinkel]

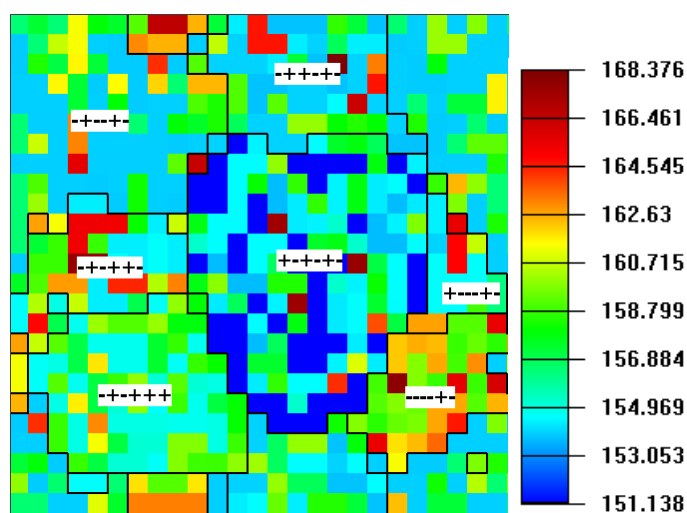
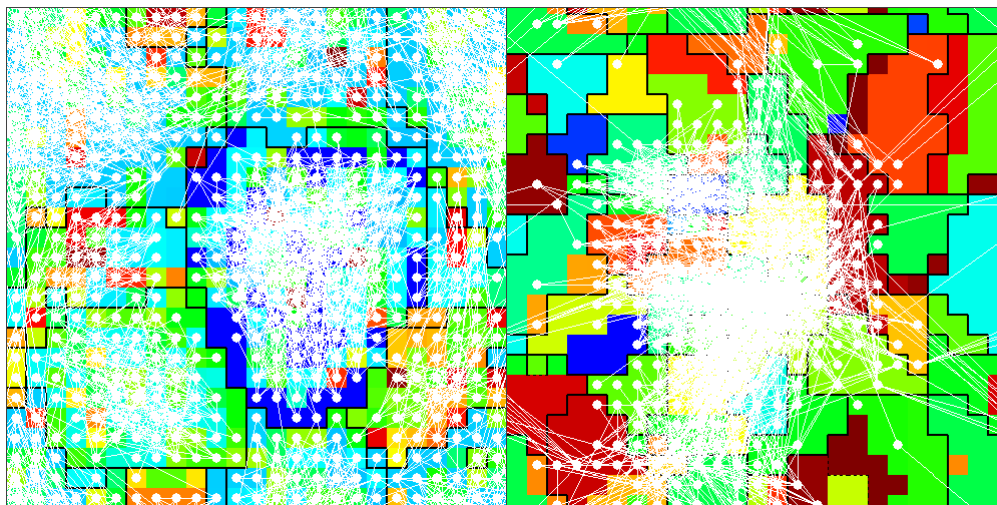


Abbildung 167: Kohonenkarte des Permethoxy-HBCs, 1100 K MD, 1ns

Auf den ersten Blick sind aus obiger Grafik im Vergleich mit Abbildung 166 zwei Dinge ersichtlich. Der durch die Moleküldynamik abgedeckte Konformationsraum der *bay*-Diederwinkel ist sehr viel kleiner als der der Konformationsuche. Außerdem existieren eine ganze Anzahl durch die Stellung der Methoxygruppe bedingter Unterkonformere, welche durch die Konformationsanalyse zuvor nur unzureichend erfasst wurden, in der Moleküldynamik aber deutlich durch unregelmäßige Färbung der Gerüstisomercluster angezeigt werden.

### 2.8.9.2.3 Konformationsräume, TRAJECTORIX-Projektion

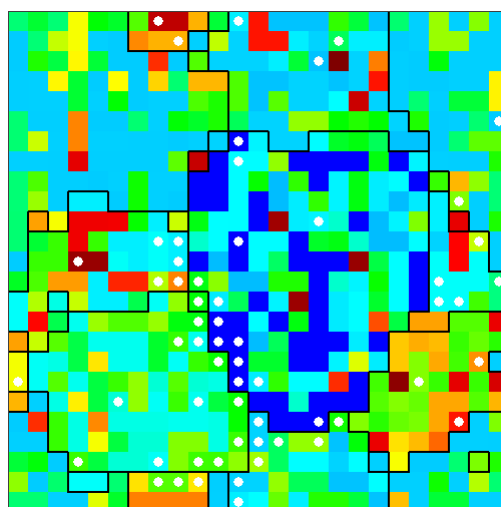
Eine präzisere Darstellung dieses Sachverhalts lässt sich wiederum durch Anwendung des TRAJECTORIX-Moduls gewinnen, siehe Abbildung 168 und Abbildung 169.



**Abbildung 168:** Projektion der 1100K MD-Trajektorie auf die Netzkarten, inklusive Verbindungslinien  
a) Netzkarte der 1100K MD-Rechnung  
b) Netzkarte der HYPERCHEM *conformational search*

In linker obiger Abbildung ist zu erkennen, dass die Übergänge aus der zentralen  $+-+ -$  Konformation sehr viel seltener sind als Übergänge zwischen anderen *bay*-Isomeren. Außerdem ist zu erkennen, dass die MD-Simulation weite Bereiche der systematischen Suche praktisch nicht erfasst.

Bei identischen Hyperflächen würde man in Abbildung 169 genau einen Punkt pro Cluster erwarten, da bei der systematischen Konformationsuche exakt eine Struktur pro Konformation erzeugt wird.



**Abbildung 169:** Projektion der HYPERCHEM *conformational search* Konformationen auf die Netzkarte der 1100 K MD

Wie ersichtlich findet sich mindestens ein Punkt pro Cluster in obiger Grafik. Dies spricht dafür, dass die Konformationsuche, zumindest was die trainierten Diederwinkel der *bay*-Region angeht, eine strikte Obermenge der MD-Simulation ist, auch wenn theoretisch nur aus dem Fehlen eines Punktes das Gegenteil zwingend gefolgert werden darf.

Andererseits sind eine Häufung von zusätzlichen Punkten an der Grenzfläche  $+--+$  zu  $-+---$  zu erkennen. Dieses spricht für Konformationen, die von der MD-Simulation nicht abgedeckt werden und deshalb „falsch“ abgebildet werden (mit großer Distanz zwischen abgebildeter Struktur und repräsentierendem Neuron).



### 3 Aromatizität von [4n]Annulen-Isomeren

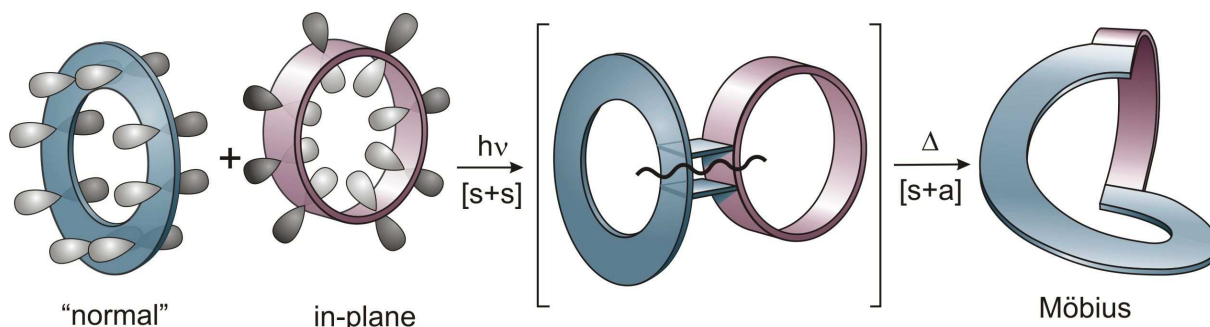
#### 3.1 Einleitung

Wie bereits im Kapitel 1.4.2 beschrieben, hängt das elektronische Energieschema und damit auch die Aromatizität bzw. Antiaromatizität von Annulenen entscheidend von der räumlichen Struktur des Moleküls ab.

Für das Vorliegen einer Möbiustopologie müssen die beteiligten Diederwinkel des zyklisch konjugierten Systems einen Phasenwechsel durchlaufen, d.h. die am  $\pi$ -System beteiligten p-Orbitale müssen insgesamt einmal um  $180^\circ$  verdrillt sein. Dies bedeutet jedoch automatisch eine Verminderung der elektronischen Konjugation, welche dem „aromatischen“ Energiegewinn entgegenwirkt. Kraftfeldprogramme zeigen für derartig verdrillte Strukturen eine erhöhte Spannungsenergie an.

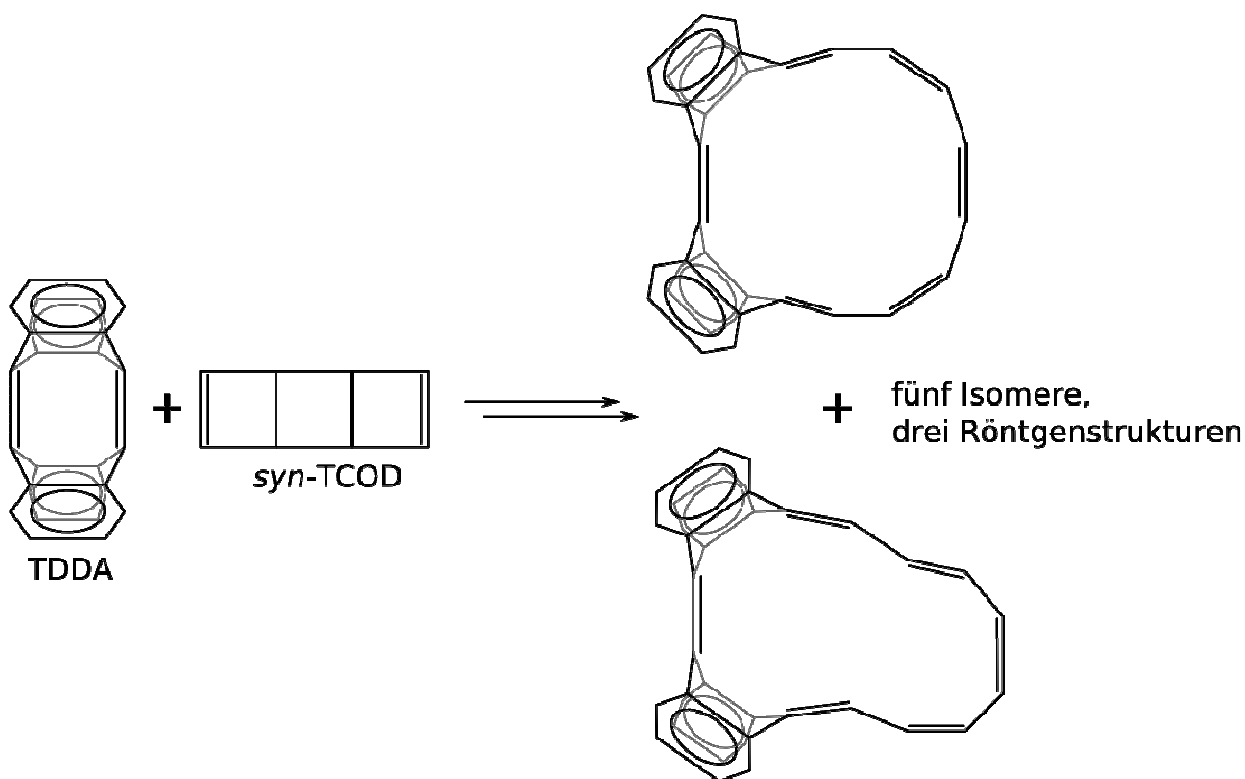
Für die unsubstituierten, neutralen [4n]Annulene mit kleiner Ringgröße gilt, dass der Energiegewinn durch Möbiusaromatizität die Spannungsenergie die durch die Verdrillung verursacht wird, nicht kompensieren kann. Selbst wenn man eine solche Verbindung in situ herstellen könnte, würde sich mit niedriger Aktivierungsbarriere "entdrillen" und in ein antiaromatisches unverdrilltes Hückel-Annulen übergehen. Das ist vermutlich der Grund warum es bis heute nicht gelungen ist ein unsubstituiertes Möbius-Annulen zu synthetisieren. In der Regel entziehen sich Hückelannulene mit  $4n$   $\pi$ -Elektronen durch Bindungslängenalternanz oder Verdrillung um  $90^\circ$  der destabilisierenden, antiaromatischen Konjugation und versuchen eine nichtaromatische elektronische Struktur anzunehmen. Das bekannteste Beispiel dürfte das Cyclooctatetraen darstellen<sup>[122]</sup>, welches nicht planar, sondern in einer Bootskonformation mit ausgeprägter Bindungslängenalternanz<sup>[123]</sup> mit C=C-(133.4pm) und C-C-(146.2pm) vorliegt. Rzepa<sup>[124]</sup> und später auch Schleyer<sup>[125]</sup> postulieren die Beteiligung von Möbiusisomeren im Isomerenequilibrium der [16]Annulene.

Die erste Synthese, Isolation und einwandfreie strukturelle Charakterisierung per Röntgenstrukturanalyse eines konjugierten organischen Möbiussystems gelang 2003 Herges, Ajami *et al.* Ein entscheidender Faktor für die gelungene Synthese dürfte die Vororientierung durch Ringmetathese aus einem „normalen“ und einem *in-plane*  $\pi$ -System bestehen, siehe Abbildung 170.



**Abbildung 170:** Metathesestrategie zur Formung eines Möbiusrings, entnommen von Herges *et al.*<sup>[84]</sup>

Bei der Synthese nach Abbildung 171, genauer beschrieben bei Herges *et al.*,<sup>[84]</sup> werden die dargestellten bianthrachinodimethan-modifizierten [16]Annulene in Möbius- und Hückelkonformationen erhalten.



**Abbildung 171:** Synthese eines bianthrachinodimethan-modifizierten [16]Annulens

Von großem Interesse in diesem Zusammenhang ist jedoch der relativ große maximale Diederwinkel  $T_{\max}$  im konjugierten Ring, der die Konjugation des Systems herabsetzt und Schleyer *et al.* in einer Kritik veranlasste, von einem nichtaromatischen System zu sprechen.<sup>[126]</sup> Diese Kritik kann aufgrund von im Rahmen dieser Arbeit erzielten und zum Teil bereits veröffentlichten<sup>[127]</sup> Ergebnissen deutlich zurückgewiesen werden. Es wird im Folgenden ein spezialisiertes und sehr sensitives Verfahren als Kombination bekannter und neuer Methoden zum Nachweis schwacher aromatischer Effekte in Annulenen entwickelt und angewandt.

### 3.2 Die KONAN-Analyse

Wie bereits in der Einleitung beschrieben, wird die Aromatizität von [4n]Annulen-Isomeren in Möbiusstruktur aufgrund des  $\pi$ -Elektronen-Orbitalschemas (siehe Kapitel 1.4.2) zwar theoretisch postuliert, eine experimentelle Darstellung und Charakterisierung eines unsubstituierten [4n]Annulens in Möbiuskonformation ist bisher aber noch nicht gelungen. Außerdem ist aufgrund der beschriebenen Verdrillung fraglich, ob sich die 4n-Möbiusannulene als Aromaten einstufen lassen. Während die Synthese und vor allem die Aufreinigung und Charakterisierung der [4n]Annulene ausgesprochen schwierig ist, sollten sich auf theoretischem Niveau Vergleiche zwischen Möbius- und Hückelisomeren sehr viel einfacher durchführen lassen. Eine systematische Analyse der Hyperfläche von [4n]Annulenen scheiterte bis jetzt aber immer am Fehlen von Methoden, die Hyperflächen dieser Moleküle vollständig zu Generieren und zu Untersuchen. Die KONAN-Analyse wird zu dem Zweck entwickelt, genau diese Methodik zu liefern und zumindest auf theoretischem Level die Frage eindeutig zu klären, ob sich für [4n]Annulene die Hückelregel umkehrt, d. h. ob Möbius-Annulene mit 4N delokalisierten Elektronen wirklich aromatisch sind. Zum anderen sollen stabile Möbiusannulene als Synthesziele identifiziert werden.

KONAN steht für „**K**onformer- und Konfigurationsisomerenaromatizität von **A**nnulenen“. Dabei handelt es sich um einen *Workflow* (Vorgehensweise und Programme), die Hyperfläche und aromatischen Eigenschaften von Annulen-Isomeren zu untersuchen, welche im Rahmen dieser Arbeit entwickelt wird. Im Falle der Annulene erschwert die große Anzahl der denkbaren und auch in der Praxis vorkommenden energetisch sehr ähnlichen Isomere die Rechnungen. Eine systematische Konformations- und Konfigurationsisomeren-suche (im Folgenden kurz: Konformations-suche) führt in direkten Konflikt mit heutzutage verfügbaren Rechenleistungen, will man die Untersuchung insgesamt auf befriedigend hohem Niveau durchführen. Folglich sind aus der Literatur noch keine systematischen Konformations-suchen mit dichtefunktionale Methoden (B3LYP/6-31G\*,<sup>[86],[87]</sup> KMLYP,<sup>[128]</sup> etc..) bekannt.

Zwei Faktoren sind maßgeblich für die Schwierigkeiten bei der Konformations-suche mit DFT- und QM-Methoden verantwortlich. Zum einen überschreitet die Anzahl der nötigen Optimierungen die verfügbare Rechenleistung. Zum anderen entstehen im Verlauf der leistungsfähigen Konformations-suchalgorithmen, die über den internen Koordinatenraum der Diederwinkel im Ring der Annulene iterieren, stark verzerrte Strukturen. Diese konvergieren

im Verlauf einer *ab-initio*, dichtefunktionalen oder semiempirischen Optimierung zu Strukturen bei denen oft unerwünschte „Nebenreaktionen“ eingetreten sind. Durch z.B. das Abspalten von Wasserstoff oder anderen Fragmenten stellt die Zielstruktur kein Annulen mehr dar und muss daher verworfen werden. Die Konformationsuche ist damit eine Domäne von Kraftfeldverfahren bei denen prinzipbedingt keine Bindungen geknüpft oder gebrochen werden und die daher die Konnektivität des Annulenrings intakt lassen. Diese können elektronische Effekte jedoch nicht mit einbeziehen und sind ungeeignet, Möbiusaromatizität zu repräsentieren.

Der Ansatz der KONAN-Analyse besteht darin, die sehr große Zahl der im Konformations- und Konfigurations-Erzeugungsschritt anfallenden Strukturen mit einem schnellen aber wenig genauen Verfahren (Kraftfeld) vorzuoptimieren und aufgrund ihrer hohen Energie unrealistischen Strukturen auszuschließen. Aus dieser bereits geringeren Anzahl von plausibleren Strukturen werden dann mit einem aufwändigeren, aber genaueren Verfahren (semiempirisch) wiederum die stabilsten ausgewählt, die dann mit einer sehr genauen, noch rechenzeitintensiveren Methode optimiert werden. D. h. um die stabilsten (und damit experimentell relevanten) Isomere eines Annulens zu finden, wird die große Menge der theoretisch denkbaren Strukturen schrittweise eingeeengt, indem auf die kleiner werdende Zahl der Strukturen zunehmend aufwendigere Rechenmethoden angewandt werden, bis eine überschaubare Anzahl der stabilsten Isomere übrigbleibt. Wichtig ist dabei, dass die ersten, ungenauen Stufen der KONAN-Analyse immer eine Übermenge an Isomeren an die folgenden Stufen weiterreichen müssen. Isomere, die in den ersten Schritten ausgeschlossen oder gar nicht erst erzeugt werden, werden von den genaueren Methoden nicht mehr erfasst. MD- und QM-Verfahren müssen also aufeinander abgestimmt verwendet werden und erlauben dann die systematische und vollständige Untersuchung von Annulen-Hyperflächen auf dichtefunktionalem Niveau.

Ein grafisches Kurzschemata dieses mehrstufigen Verfahrens wird in Abbildung 172 vorgestellt.

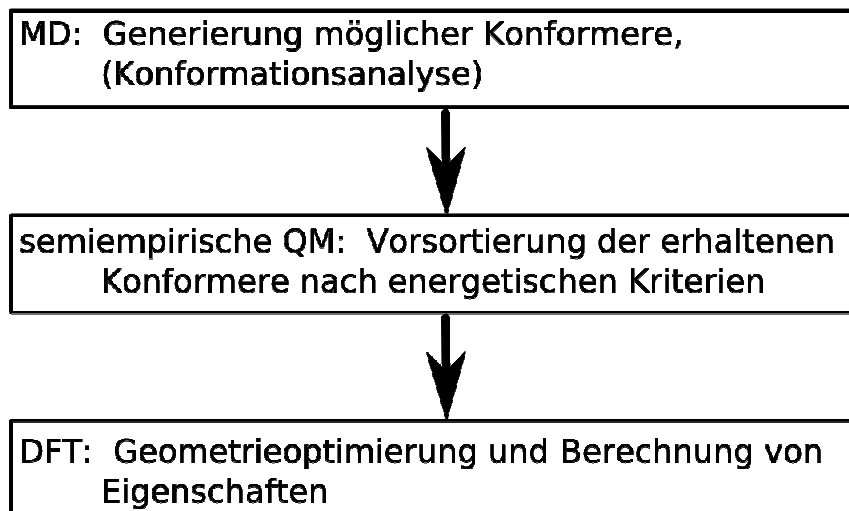


Abbildung 172: Kurzschemata einer KONAN-Analyse

### 3.3 CANE-Deskriptoren

In [n]Annulenen gibt es n konnektivitätsäquivalente Kohlenstoffzentren. Die gleiche Struktur, die sich nur in der internen Nummerierung äquivalenter Atome unterscheidet, kann N-mal gefunden werden, wie anhand Abbildung 173 dargestellt ist. Existierende IUPAC-Nomenklatorschemata sind für die Unterscheidung der Konformations- und Konfigurationsisomeren nur sehr begrenzt von Nutzen.

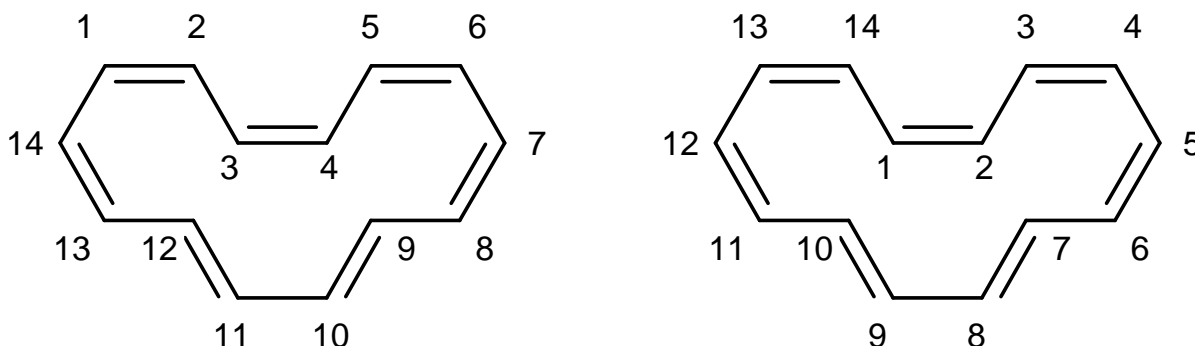


Abbildung 173: zwei mögliche Numerierungsisomere des 14-Annulens

Die meisten gebräuchlichen Programme wie zum Beispiel HYPERCHEM und GAUSSIAN erkennen eine solche Äquivalenz zweier Strukturen nicht. Automatisiert sind diese Strukturen nicht trivial als identisch zu erkennen. Es wird an dieser Stelle speziell für die Identifikation von derartigen Isomeren das CANE-Nomenklatorschema eingeführt (*Canonical Annulene Enumeration*). Die Berechnung ist zusammen mit verschiedenen, bereits im Rahmen der Einleitung, Kapitel 1.4, beschriebenen, geläufigen Aromatizitätskriterien im Programm XYZ2CANE.EXE, Kapitel 5.3.21, implementiert.

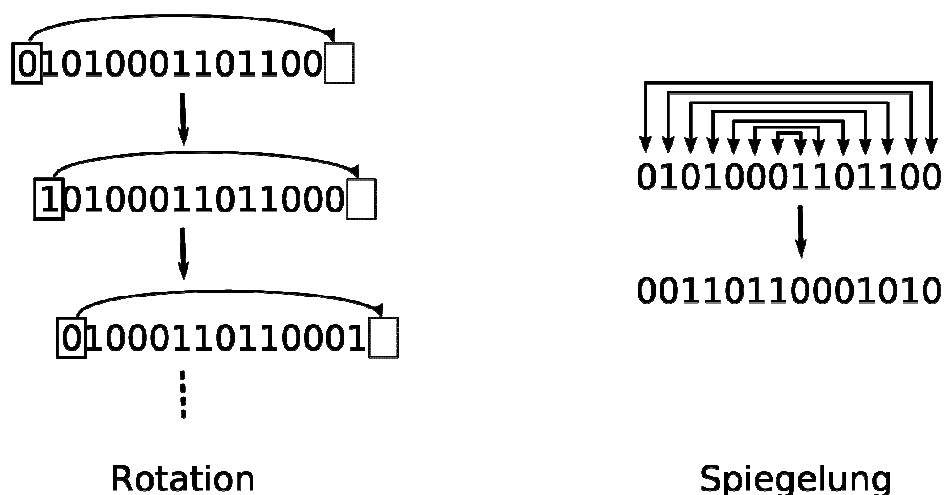
Im Rahmen des CANE Nomenklatorschemas werden die Bindungen zwischen den Kohlenstoffatomen unterteilt in *cis*(0)- für Diederwinkel ( $-90^\circ$  bis  $90^\circ$ ) und

*trans*(1)-Bindungen ( $90^\circ$  bis  $180^\circ$  und  $-90^\circ$  bis  $-180^\circ$ ), unabhängig davon, ob es sich um Einzel- oder Doppelbindungen handelt. Probleme mit der Klassifikation der Bindungen (Bindungslängenausgleich), wie sie in konjugierten Doppelbindungen sonst regelmäßig auftauchen, werden damit vermieden. Von einem beliebigen Startatom ausgehend werden ringförmig sämtliche Bindungen zwischen benachbarten Kohlenstoffatomen aufgezählt. Aus Abbildung 173 ergeben sich damit folgende Sequenzen:

links: 01010001101100

rechts: 01000110110001

Wird diese Sequenz als binäre Zahl interpretiert, ist sofort offensichtlich, dass Numerationsisomere durch Rotationen (von denen es insgesamt  $N$  gibt) ineinander überführt werden können, siehe Abbildung 174.



**Abbildung 174:** Rotations- und Spiegeloperationen zur Überführung von Numerationsisomeren

Eine weitere Möglichkeit der Numerationsisomerie besteht in der beliebigen Richtung der Aufzählung. Geometrisch und in der Darstellung als Binärzahl entspricht dies einer Spiegelung des Moleküls. Damit ergeben sich insgesamt maximal  $2N$  äquivalente Sequenzen. Als CANE-Deskriptor eines Annulens ist diejenige Darstellung definiert, welche der kleinsten Zahl aller denkbaren Spiegelungs- und Rotationsvarianten entspricht. Im obigen Beispiel aus Abbildung 173 ist dies 00010100011011.

Äquivalente Strukturen werden damit garantiert durch denselben CANE-Deskriptor repräsentiert, was die Eliminierung von Dubletten ermöglicht. Gerade bei größeren Isomeren können auch nichtäquivalente Strukturen gleiche CANE-Deskriptoren erhalten. Diese sind aber in der Regel schon nach

der Kraftfeldvoroptimierung durch unterschiedliche Energien zu unterscheiden.

Die maximale Zahl an unterschiedlichen CANE-Deskriptoren kann einfach durch Abzählen aller möglichen Varianten gewonnen werden, für die ersten 32 Annulene mit Hilfe des Programms EXAKTE\_LOESUNG, Kapitel 5.3.7, durchgeführt in Tabelle 73. Die obere Grenze kann zu  $2^N$  abgeschätzt werden, welche einen einfacheren Ansatz ohne CANE-Isomerieerkennung repräsentiert. Der Vergleich zeigt eindrucksvoll den Nutzen.

**Tabelle 73: CANE-Deskriptoren bis zum [32]Annulen**

Anzahl C-Atome (N)	CANE-Isomere exakt	Obere Grenze $2^N$
4	6	16
5	8	32
6	13	64
7	18	128
8	30	256
9	46	512
10	78	1024
11	126	2048
12	224	4096
13	380	8192
14	687	16384
15	1224	32768
16	2250	65536
17	4112	131072
18	7685	262144
19	14310	524288
20	27012	1048576
21	50964	2097152
22	96909	4194304
23	184410	8388608
24	352698	16777216
25	675188	33554432
26	1296858	67108864
27	2493726	134217728
28	4806078	268435456
29	9272780	536870912
30	17920860	1073741824
31	34669602	2147483648
32	67159050	4294967296

### 3.4 Übersicht der KONAN-Analyse

Im Wesentlichen zusammengefasst wird der Datenfluss der KONAN-Analyse durch Abbildung 175.

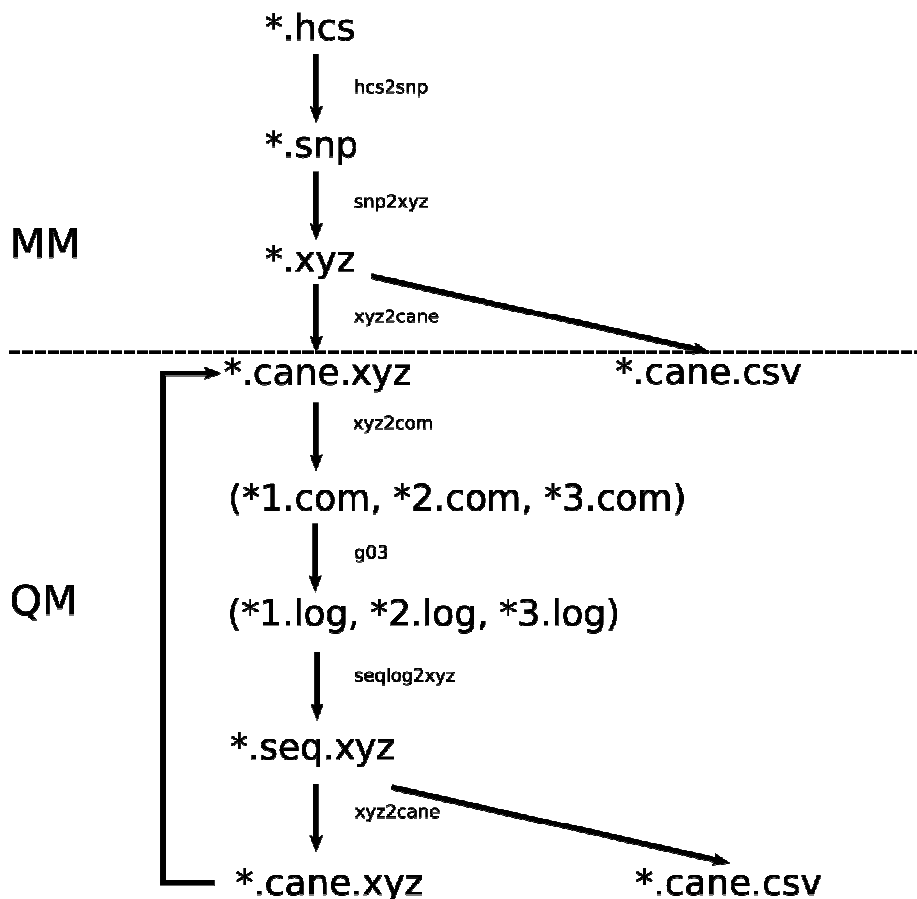


Abbildung 175: vereinfachter Datenfluss im Verlauf der KONAN-Analyse

Die obige Abbildung zeigt die zu verwendenden Programme und die bei jedem Schritt erhaltenen Dateitypen. Eine präzisere Übersicht der KONAN-Analyse ist in Abbildung 176 dargestellt. Die erhaltenen Isomerendatensätze werden durch Operationen erhalten, welche als Pfeile repräsentiert werden. Die `*.csv`-Dateien enthalten eine Auflistung der Isomere, ihrer Energien und aller bis dato berechneten Aromatizitätsdeskriptoren. Als *comma separated values* liegen diese damit in einem Format vor, das von allen gängigen Tabellenkalkulationen und Statistikpaketen zur weiteren Verarbeitung direkt gelesen werden kann. Im Folgenden werden die einzelnen Operationen inklusive der verwendeten Programme und Kommandozeilenparameter im Detail anhand des Beispiels [12]Annulen beschrieben.



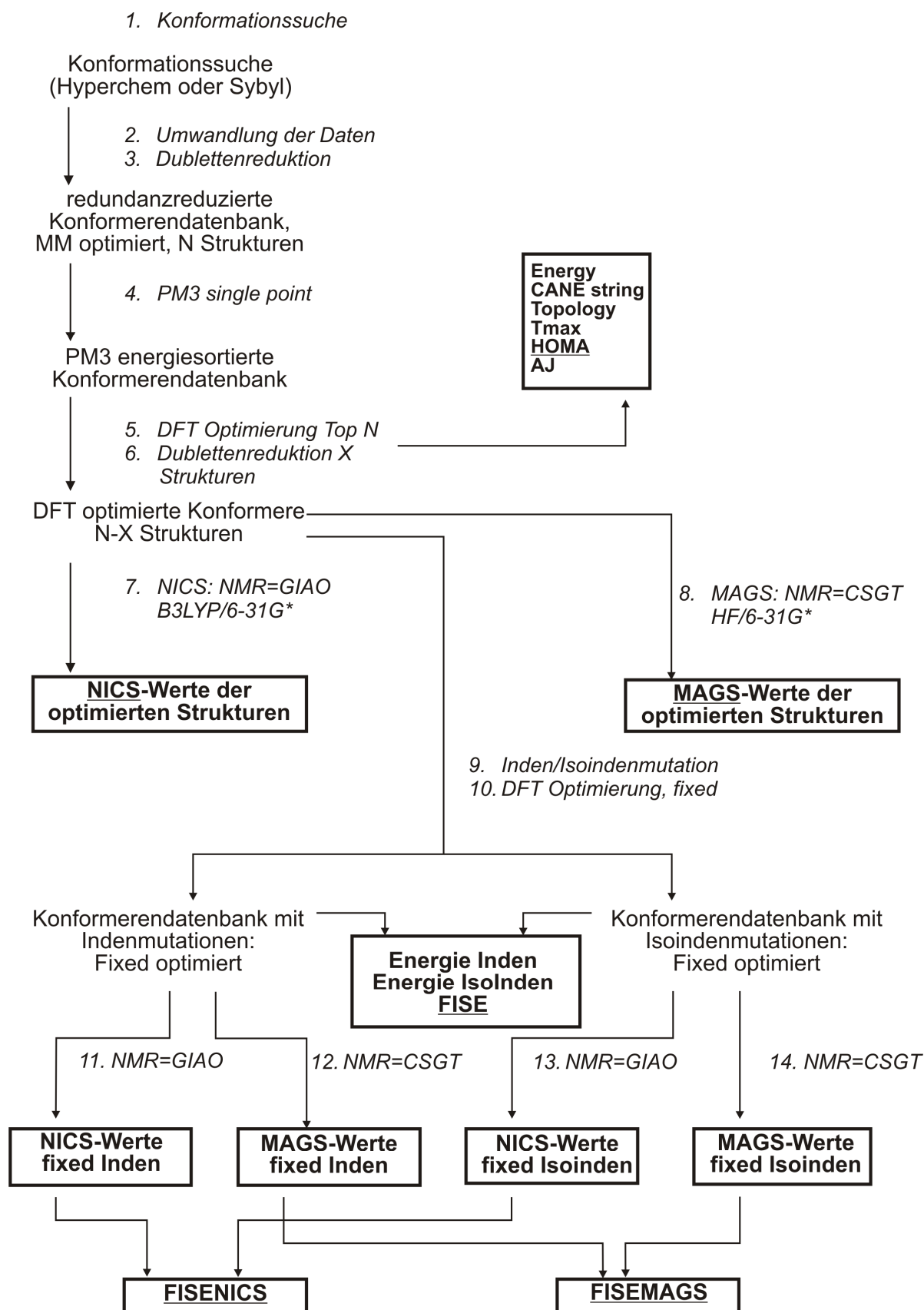


Abbildung 176: KONAN-Flieschema

### 3.4.1 Konformationssuche

Die Konformations- und Konfigurationsisomerensuche wird mit Hilfe externer Kraftfeldprogramme durchgefhrt; die erhaltenen Daten werden ins xyz-

Dateiformat überführt, welches im weiteren Verlauf die Rolle als Datensatzformat übernimmt.

In Vorversuchen wurden die *conformational search engines* der Programme HYPERCHEM 6.03<sup>[37]</sup> mit den Kraftfeldern MM+ und OPLS, SYBYL 6.9<sup>[39]</sup> (DISCOsearch) mit dem Kraftfeld MMFF94<sup>[35]</sup> getestet.

Als ungeeignet erwies sich die Kombination HYPERCHEM + OPLS, da die erhaltenen Strukturen schlechte Startstrukturen für die DFT-Geometrieoptimierung darstellen. Sowohl SYBYL/MMFF94 als auch HYPERCHEM/MM+ ergeben vergleichbare Ergebnisse nach Anzahl und Qualität der erhaltenen Isomeren. Die wesentlich einfachere Handhabung und der rechenzeiteffizientere Algorithmus des Programms HYPERCHEM geben zunächst den Ausschlag für die Wahl als Standardmethode in dieser Arbeit.

### 3.4.2 Verzeichnisstruktur

Eine Konformationsuche ergibt große Mengen an Dateien als Ergebnisse und Zwischenergebnisse. Die Organisation dieser Daten auf dem Datenträger ist wissenschaftlich irrelevant, soll aber für nachfolgende Mitarbeiter dokumentiert werden. Jedem Annulen ist ein Stammverzeichnis zugeordnet. Die Zwischenergebnisse einzelner Arbeitsgänge stehen in Unterordnern in diesem Stammverzeichnis.

Die Stammverzeichnisnamen einer Konformationsuche werden nach Anzahl der Ringe und Ladung benannt.

**Tabelle 74: Stammverzeichnisnamen für die KONAN-Analyse**

Ladung	Verzeichnisname generisch (X: Anzahl der Ringatome)	Verzeichnisname Beispiele
neutral	X	16
kationisch (+)	Xp	17p, 13p
anionisch(-)	Xm	15m, 11m
kationisch(2+)	Xpp	18pp
anionisch(2-)	Xmm	14mm

Im Rahmen der KONAN-Analyse stellt die Konformationsuche als erster Schritt die Weichen für den gesamten späteren Verlauf. Strukturen, die an dieser Stelle nicht generiert werden, sind für die Analyse „verloren“, da in den folgenden Operationen nur noch Strukturen eliminiert werden und keine neuen dazukommen. Auf einer Konformationsuche aufbauende Rechnungen werden in einem Unterordner, benannt nach der Suche, untergebracht. So

steht z.B. „mmp“ für die Kombination *HYPERCHEM Conformational Search* + Kraftfeld MM+. „disco“ steht für die Kombination *Sybyl/disco* + MMFF94.

Die verwendeten Methoden und Basissätze der weiteren *single-point* Rechnungen und Optimierungen sind frei wählbar. Es wird folgende Unterordnerstruktur in Tabelle 75 vorgeschlagen:

**Tabelle 75: Unterverzeichnisse in einem *conformational-search* Ordner :**

Unterordner	Beschreibung
sel	Definition der Diederwinkel für das xyz2cane Programm, <i>selection files</i>
cs	Konformationsuche
pm3	PM3 <i>single-point</i> Rechnungen
kmlyp_6-31Gd	KMLYP/6-31G* Rechnungen
b3lyp_6-31Gd	B3LYP/6-31G* Rechnungen
...	Weitere QM Niveaus

**Tabelle 76: Unterverzeichnisse in einem Methodenordner, z.B. b3lyp\_6-31Gd**

Unterordner	Beschreibung
opt	Optimierung
mags	magnetic susceptibility: ( <i>ab initio</i> ) HF/6-31G* NMR=csgt %subst L1002 /usr/local/g03/original
nics	NMR-Rechnung
opt_finden	fixed indene Optimierung
opt_fisoinden	fixed isoindene Optimierung
nics_finden	NMR-Rechnung für die optimierten <i>fixed indenenes</i>
nics_fisoinden	NMR-Rechnung für die optimierten <i>fixed isoindenenes</i>
mags_finden	<i>magnetic susceptibility</i> der optimierten <i>fixed indenenes</i> : HF/6-31g* NMR=CSGT %subst l1002 /usr/local/g03/original
mags_fisoinden	<i>magnetic susceptibility</i> der optimierten <i>fixed isoindenenes</i> : HF/6-31g* NMR=CSGT %subst l1002 /usr/local/g03/original

Eine Baumstruktur der Verzeichnisse für die KONAN-Analyse eines [12]Annulens ist in Abbildung 177 wiedergegeben.

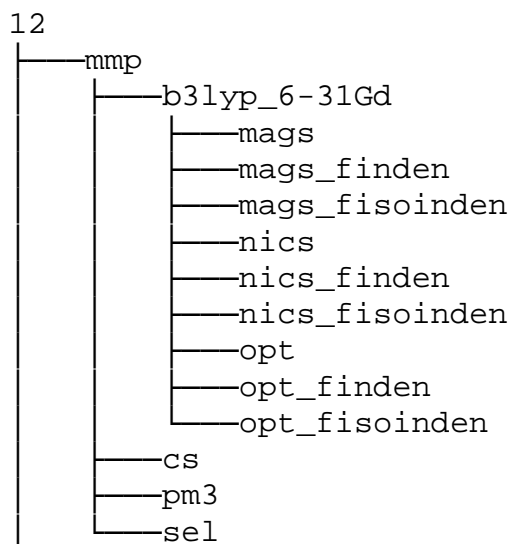


Abbildung 177: Verzeichnisstruktur [12]Annulen

### 3.5 [12]Annulen

Das [12]Annulen wurde zuerst von Schröder et al. synthetisiert und die konformativen Eigenschaften wurden durch NMR-Studien untersucht.<sup>[129]</sup> Anhand des [12]Annulens soll beispielhaft eine schrittweise Beschreibung der verwendeten Programme nach dem Schema aus Abbildung 176 erfolgen. Der sich ergebende Verzeichnisbaum ist bereits in Abbildung 177 wiedergegeben und der Arbeit auf DVD beigelegt im *supplementary material*.

#### 3.5.1 Konformationssuche

Die Konformationssuche wird durchgeführt mit dem Programm „HYPERCHEM / Conformational Search / MM+“ unter Variation sämtlicher Ringdiederwinkel. Das Molekül wird mit alternierenden Doppel- und Einfachbindungen modelliert, da die Darstellung als aromatischer Ring mit Eineinhalbfachbindungen zu schlechten Startstrukturen für die quantenmechanische Optimierung führt. Erhalten werden im Verzeichnis 12/mmp/cs die Dateien c12\_mmp\_csearch.HCS mit den gefundenen Isomeren und c12\_mmp\_csearch.hin mit einer Beschreibung des Moleküls inklusive der Ringdiederwinkel.

#### 3.5.2 Umwandlung der Daten, Dublettenreduktion

Die Isomeren aus der hcs-Datei werden in das snp-Format und von dort ins xyz-Format überführt mit Hilfe der Befehle.

```

12/mmp/cs $ generate_snap_headers -sc12_mmp_csearch.snp -
            ic12_mmp_csearch.hin

12/mmp/cs $ hcs2snp c12_mmp_csearch.HCS
            c12_mmp_csearch.snp

```

```
12/mmp/cs$ snp2xyz -p c12_mmp_csearch.snp
c12_mmp_csearch.xyz
```

Die Ringtorsionen können aus den *selections* der hin-Datei erhalten werden:

```
12/mmp/cs $ hin2sel.bat -i c12_mmp_csearch.hin >
../sel/c12ringtors.sel
```

Eine Dublettenreduktion mit Hilfe der CANE-Deskriptoren ergibt die Dateien `c12_mmp_csearch.cane.xyz` und `c12_mmp_csearch.cane.csv`

```
12/mmp $ xyz2cane -s sel/c12ringtors.sel -i
cs/c12_mmp_csearch.xyz -o c12_mmp_csearch.cane.xyz -v
> c12_mmp_csearch.cane.csv
```

Damit ist der MM-Teil aus Abbildung 175 abgeschlossen. Ein Ausschnitt der erhaltenen Datei `c12_mmp_csearch.cane.csv` ist in Tabelle 77 wiedergegeben. Von den insgesamt erzeugten 9833 Strukturen werden ca. 90% verworfen.

**Tabelle 77:** [12]Annulen - Die energetisch stabilsten 25 Isomere (MM+ Energie), vollständige Tabelle im Zusatzmaterial auf DVD

lognr	Erel(kcal/mol)	CANE string	Topology	HOMA
1	0	000011000011	HUECKEL	1.4597
13	6.6296	000001100011	HUECKEL	1.4514
37	7.3511	001100110011	HUECKEL	1.3097
61	7.7385	000001000001	HUECKEL	1.4327
73	7.8578	000110011011	HUECKEL	1.3368
121	7.9869	000110001111	HUECKEL	1.373
145	10.113	001100110011	HUECKEL	1.2778
146	12.4008	000010100011	HUECKEL	1.3965
194	20.797	000010000111	HUECKEL	1.3927
218	21.1205	000011001111	HUECKEL	1.3724
266	22.3745	000110011011	HUECKEL	1.2795
313	22.683	000001000011	MOEBIUS	1.424
358	26.0433	000000110011	HUECKEL	1.4198
381	27.4449	000110011011	HUECKEL	1.3471
402	28.6518	000000100001	HUECKEL	1.4216
421	29.6533	000001000011	MOEBIUS	1.4213
464	30.6217	000000100011	MOEBIUS	1.4121
511	30.7405	000110011011	HUECKEL	1.3578
559	31.2591	000011000011	HUECKEL	1.3759
568	31.7819	000110010011	MOEBIUS	1.3714
592	31.8711	000110001111	HUECKEL	1.3283
615	32.2944	001100110011	HUECKEL	1.2882
639	32.8571	000011000111	MOEBIUS	1.41
674	32.8574	000111000111	HUECKEL	1.3628
685	33.089	001100110011	HUECKEL	1.2998
Total numbers of structures processed				9833
Number of structures written to output				1065
Number of Doublettes: (CANE and energy same)				8768
Total Number of CANE - doublettes: (same CANE different energy)				951

Es zeigt sich eine Dominanz der Hückelstrukturen bei den Isomeren mit niedriger Kraftfeldenergie. Als Anmerkung sei erwähnt, dass aus Platzgründen einige Spalten der csv-Datei nicht aufgeführt werden, diese sind jedoch in den Originaldateien im *supplementary material* auf der DVD zu finden. Die Bedeutung der Spaltenköpfe ist in Tabelle 78 aufgeführt.

**Tabelle 78: Spaltenköpfe der KONAN csv-Dateien**

Tabellenkopf	Beschreibung
structurenr	Laufende Nummer der Struktur in der Eingabe xyz-Datei
lognumber	Laufende Nummer des Isomers in der ursprünglichen Konformationssuche
E(kcal/mol)	Energie in kcal/mol
Erel(kcal/mol)	Relative Energie bezogen auf das energetisch niedrigste Isomer in dieser Tabelle
CANE string	CANE Deskriptor
Topology	Topologie (HUECKEL oder MOEBIUS)
dmax	Maximale zentrale Bindungslänge der betrachteten Diederwinkel (definiert in der .sel Datei)
dmin	Minimale zentrale Bindungslänge der betrachteten Diederwinkel (definiert in der .sel Datei)
ddiff	Differenz der obigen Werte
Tmax	Größte Abweichung eines der betrachteten Diederwinkel aus der Planarität (0° und 180°)
Rav	Durchschnittliche Bindungslänge der zentralen Bindungen der Torsionswinkel
f	Parameter f nach Gleichung (1.37)
EN	Strukturelles Aromatizitätskriterium nach Gleichung (1.36)
GEO	Strukturelles Aromatizitätskriterium nach Gleichung (1.38)
HOMA	Strukturelles Aromatizitätskriterium nach Gleichung (1.34)
A <sub>j</sub>	Julg-Index, strukturelles Aromatizitätskriterium nach Gleichung (1.33)
NICS	-NICS(0), magnetisches Aromatizitätskriterium mit umgekehrtem Vorzeichen zur Literatur
MAGS	<i>magnetic susceptibility</i> , magnetisches Aromatizitätskriterium
indene energy	Energie der ( <i>fixed</i> ) <i>indenes</i> , siehe Kapitel 1.4.3
isoindene energy	Energie der ( <i>fixed</i> ) <i>isoindenes</i> , siehe Kapitel 1.4.3
ISE	(F)ISE-Wert nach Gleichung (1.27)

### 3.5.3 PM3 single-point

Diese Kraftfeldenergie berücksichtigt elektronische Aromatizitätseffekte prinzipiell nicht. Zwecks realistischerer relativer Energien wird auf jeder der 1065 nach Dublettenreduktion übrig gebliebenen Strukturen eine semiempirische *single-point* Rechnung durchgeführt. Die Vorgehensweise verläuft ab jetzt gemäß der QM - Schleife in Abbildung 175.

```
12/mmp $ cp c12_mmp_csearch.cane.xyz pm3/c12_mmp_pm3.xyz
12/mmp/pm3 $ xyz2com -k "pm3" -c 0 -m 1 -g g98 -s
c12_mmp_pm3.xyz
```

Dies erzeugt eine Sequenz an GAUSSIAN *input files*, welche durch die ebenfalls erzeugte .bat Datei (bzw. .sh unter LINUX) nacheinander abgearbeitet werden.

```
12/mmp/pm3 $ c12_mmp_pm3.bat
(bzw. sh c12_mmp_pm3.sh auf Un*x Maschinen.)
```

Die Sequenz der erhaltenen GAUSSIAN *output* Dateien wird in eine xyz-Isomerendatenbank überführt mittels des Programms SEQLOG2XYZ.

```
12/mmp/pm3 $ seqlog2xyz -s .log -a 1 -z 20000 c12_mmp_pm3.
> c12_mmp_pm3.seq.xyz
```

Der Aufruf von XYZ2CANE ergibt die Dateien c12\_mmp\_pm3.cane.xyz und c12\_mmp\_pm3.cane.csv.

```
12/mmp $ xyz2cane -s sel/c12ringtors.sel -i
pm3/c12_mmp_pm3.seq.xyz -o c12_mmp_pm3.cane.xyz -v -c
> c12_mmp_pm3.cane.csv
```

Sortieren nach relativer Energie ergibt eine neue Reihenfolge und im Durchschnitt eine Begünstigung der Isomere in Möbiustopologie, siehe Tabelle 79.

**Tabelle 79:** [12]Annulen - Die energetisch stabilsten 25 Isomere (MM+//PM3, PM3 *single-point* Energie für die auf Kraftfeldniveau optimierten Strukturen), vollständige Tabelle im Zusatzmaterial auf DVD

lognumber	Erel(kcal/mol)	CANE string	Topology	HOMA
37	0	001100110011	HUECKEL	1.3097
313	1.7014	000001000011	MOEBIUS	1.4237
145	3.1256	001100110011	HUECKEL	1.2777
73	3.7654	000110011011	HUECKEL	1.3366
421	3.9989	000001000011	MOEBIUS	1.4215
941	6.8289	000100110011	MOEBIUS	1.3325
61	8.2232	000001000001	HUECKEL	1.4328
146	8.8251	000010100011	HUECKEL	1.3963
894	9.6102	000100110011	MOEBIUS	1.3462
464	9.7312	000000100011	MOEBIUS	1.412
13	10.3388	000001100011	HUECKEL	1.4517
1	11.8616	000011000011	HUECKEL	1.4597
639	12.8266	000011000111	MOEBIUS	1.4101
1454	13.8363	000110010011	MOEBIUS	1.3571
1738	14.4249	000100011011	MOEBIUS	1.3018
568	15.3234	000110010011	MOEBIUS	1.3718
194	15.5487	000010000111	HUECKEL	1.3931
266	16.3103	000110011011	HUECKEL	1.2796
674	16.6252	000111000111	HUECKEL	1.3631
2571	18.3373	000100011011	MOEBIUS	1.2934
2489	18.9842	001101100111	MOEBIUS	1.2847
121	19.5837	000110001111	HUECKEL	1.3729
2588	21.8501	001101100111	MOEBIUS	1.2848
2370	23.3307	001101100111	MOEBIUS	1.2919
3130	23.3622	000010011011	MOEBIUS	1.3001
Total numbers of structures processed				1065
Number of structures written to output				1044
Number of Doublettes: (CANE and energy same)				21
Total Number of CANE - doublettes: (same CANE, different energy)				930

### 3.5.4 DFT-Optimierung

Die DFT-Optimierung transformiert die Startstrukturen aus der Kraftfeldoptimierung in die Koordinaten, für welche die Aromatizitätsdeskriptoren berechnet werden. In diesem Fall wird auf dem Niveau B3LYP/6-31G\*[86],[87] gearbeitet.

```
12/mmp $ cp c12_mmp_pm3.cane.xyz b3lyp_6-
31Gd/opt/c12_mmp_opt_b3lyp_6-31Gd.xyz

12/mmp/b3lyp_6-31Gd/opt $
xyz2com.bat -k "b3lyp/6-31G* opt freq=noraman" -c 0 -
m 1 -g g03 -s c12_mmp_opt_b3lyp_6-31Gd.xyz
```

In der Regel interessieren nur die Isomere mit niedrigen Energien, weshalb an dieser Stelle die Zahl der zu optimierenden Isomere mit Hilfe eines *energy*



*cutoffs* reduziert werden kann. Im konkreten Fall werden nur Isomere mit einer relativen MM+//PM3-Energie von weniger als 50 kcal/mol zur Optimierung herangezogen. Dies sind 72 Isomere, von denen nur die 25 energetisch niedrigsten in Tabelle 79 wiedergegeben werden. Eine Skriptdatei zur Optimierung dieser Isomere kann aus der lognumber Spalte der vollständigen energiesortierten PM3-Tabelle im Zusatzmaterial einfach erzeugt werden, hier unter dem Namen `c12_mmp_opt_b3lyp_6-31Gd.sh`.

```
12/mmp/b3lyp_6-31Gd/opt $  
sh c12_mmp_opt_b3lyp_6-31Gd.sh
```

Nach Überprüfung der Ergebnisse auf Korrektheit und eventueller manueller Nachoptimierung wird die log-Dateiensequenz wieder in eine xyz-Datei überführt und per `xyz2cane` ausgewertet.

```
12/mmp/b3lyp_6-31Gd/opt $  
seqlog2xyz -s .log -a 1 -z 10000 c12_mmp_opt_b3lyp_6-  
31Gd. > c12_mmp_opt_b3lyp_6-31Gd.seq.xyz  
  
12/mmp/ $ xyz2cane -s sel/c12ringtors.sel -i b3lyp_6-  
31Gd/opt/c12_mmp_opt_b3lyp_6-31Gd.seq.xyz -o  
c12_mmp_opt_b3lyp_6-31Gd.cane.xyz -v -c >  
c12_mmp_opt_b3lyp_6-31Gd.cane.csv
```

Nach Geometrieoptimierung auf DFT-Niveau und Dublettenreduktion verbleiben von den 72 Isomeren mit der niedrigsten PM3-Energie nur noch 39 im Arbeitssatz, siehe Tabelle 80. Bei den in der Tabelle angegebenen Werte für Erel handelt es sich natürlich um die relativen DFT-Energien (B3LYP/6-31G\*)

Tabelle 80: [12]Annulen – DFT-optimierte Strukturen (B3LYP/6-31G\*)

lognumber	Erel(kcal/mol)	CANE string	Topology	HOMA
146	0	000010100011	HUECKEL	0.1267
61	1.6282	000001000001	HUECKEL	0.2644
37	1.8557	001100110011	HUECKEL	0.0446
639	5.3723	000111000111	HUECKEL	-0.1307
13	5.4299	000001100011	HUECKEL	-0.0115
674	5.4753	000111000111	HUECKEL	-0.0657
3327	6.1563	000001100011	HUECKEL	0.0077
421	6.2518	000001000011	MOEBIUS	0.6999
1	6.6838	000011000011	HUECKEL	0.1852
194	6.7346	000010000111	HUECKEL	-0.0142
1454	7.3672	000110010011	MOEBIUS	-0.053
1212	8.8988	000000100011	MOEBIUS	-0.0019
121	9.7758	000110001111	HUECKEL	0.0987
358	10.4094	000100110011	MOEBIUS	-0.0353
568	11.007	000110010011	MOEBIUS	-0.0238
1007	11.1253	000110011011	HUECKEL	0.1103
1738	11.375	000100011011	MOEBIUS	-0.1427
2489	12.2926	001101100111	MOEBIUS	-0.1766
266	13.8368	000110011011	HUECKEL	0.0777
6513	14.5144	000100010001	MOEBIUS	-0.3754
2960	15.2592	000100110011	MOEBIUS	0.1938
2880	16.282	000000010001	HUECKEL	-0.1947
2896	17.4955	000010011001	HUECKEL	-0.1557
2916	18.1087	000101000101	HUECKEL	-0.3606
218	18.8539	000011001111	HUECKEL	0.0851
Total numbers of structures processed				72
Number of structures written to output				39
Number of Doublettes: (CANE and energy same)				32
Total Number of CANE - doublettes: (same CANE, different energy)				9

### 3.5.5 NICS-Rechnungen

Die NICS(0)-Rechnungen,<sup>[73]</sup> siehe Kapitel 1.4.4, setzen ein *ghost-atom* im Zentrum eines planaren Rings voraus. Dies ist im Falle der nicht planaren Annulene nicht möglich. Stattdessen wird das geometrische Zentrum aller im *selection file* aufgeführten Atome verwendet und mit Hilfe des Programms XYZ2CANE in einer xyz-Datei gespeichert.

```
12/mmp/ $ xyz2cane -s sel/c12ringtors.sel -i
c12_mmp_opt_b3lyp_6-31Gd.cane.xyz -c -n b3lyp_6-
31Gd/nics/c12_mmp_nics_b3lyp_6-31Gd.xyz
```

Nach dem Schema aus Abbildung 175 werden die GAUSSIAN *input files* erzeugt und gestartet.

```
12/mmp/b3lyp_6-31Gd/nics $
xyz2com.bat -k 'b3lyp/6-31G* NMR=GIAO' -c 0 -m 1 -g
g03 -s c12_mmp_nics_b3lyp_6-31Gd.xyz
```

```
12/mmp/b3lyp_6-31Gd/nics $  
    nohup sh c12_mmp_nics_b3lyp_6-31Gd.sh &
```

### 3.5.6 MAGS-Rechnungen

Die magnetische Suszeptibilität kann direkt aus einer NMR=CSGT<sup>[130]</sup> GAUSSIAN Rechnung bestimmt werden, es wird an dieser Stelle eine ab initio anstatt einer dichtefunktionalen Methode verwendet.

```
12/mmp $ cp c12_mmp_opt_b3lyp_6-31Gd.cane.xyz b3lyp_6-  
    31Gd/mags/c12_mmp_mags_b3lyp_6-31Gd.xyz  
  
12/mmp/b3lyp_6-31Gd/mags $  
    xyz2com.bat -k 'hf/6-31G* NMR=CSGT' -c 0 -m 1 -g g03  
    -s c12_mmp_mags_b3lyp_6-31Gd.xyz  
  
12/mmp/b3lyp_6-31Gd/mags $  
    nohup sh c12_mmp_mags_b3lyp_6-31Gd.sh &
```

NICS- und MAGS-Werte können jetzt zusammen mit den Daten aus der vorhergehenden Optimierung in eine xyz-Datei überführt werden.

```
12/mmp/$  
    seqlog2xyz b3lyp_6-31Gd/opt/c12_mmp_opt_b3lyp_6-31Gd.  
    -m b3lyp_6-31Gd/mags/c12_mmp_mags_b3lyp_6-31Gd. -n  
    b3lyp_6-31Gd/nics/c12_mmp_nics_b3lyp_6-31Gd. >  
    c12_mmp_nicsmags_b3lyp_6-31Gd.seq.xyz  
  
12/mmp/ $  
    xyz2cane -s sel/c12ringtors.sel -i  
    c12_mmp_nicsmags_b3lyp_6-31Gd.seq.xyz -o  
    c12_mmp_nicsmags_b3lyp_6-31Gd.cane.xyz -v -c >  
    c12_mmp_nicsmags_b3lyp_6-31Gd.cane.csv
```

Der Aufruf von XYZ2CANE bewirkt natürlich keine Dublettenreduktion mehr, sondern extrahiert lediglich die neuen Aromatizitätsdeskriptoren in die Datei `c12_mmp_nicsmags_b3lyp_6-31Gd.cane.csv`, eingetragen in Tabelle 81:

Tabelle 81: [12]Annulen - magnetische Aromatizitätskriterien. (DFT)

lognr	Erel (kcal/mol)	CANE string	Topology	HOMA	NICS*	MAGS
146	0	000010100011	HUECKEL	0.1267	-6.21	-72.2936
61	1.6282	000001000001	HUECKEL	0.2644	-6.95	-75.4626
37	1.8557	001100110011	HUECKEL	0.0446	3.39	-60.9554
639	5.3723	000111000111	HUECKEL	-0.1307	-4.53	-69.6015
13	5.4299	000001100011	HUECKEL	-0.0115	-2.89	-67.9704
674	5.4753	000111000111	HUECKEL	-0.0657	-4.18	-68.3812
3327	6.1563	000001100011	HUECKEL	0.0077	-2.53	-72.3696
421	6.2518	000001000011	MOEBIUS	0.6999	-14.69	-95.6884
1	6.6838	000011000011	HUECKEL	0.1852	4.17	-48.1591
194	6.7346	000010000111	HUECKEL	-0.0142	-2.02	-65.3254
1454	7.3672	000110010011	MOEBIUS	-0.053	0.04	-65.3223
1212	8.8988	000000100011	MOEBIUS	-0.0019	-2.32	-70.3135
121	9.7758	000110001111	HUECKEL	0.0987	-0.14	-58.396
358	10.4094	000100110011	MOEBIUS	-0.0353	-1.66	-71.4554
568	11.007	000110010011	MOEBIUS	-0.0238	0.23	-65.7926
1007	11.1253	000110011011	HUECKEL	0.1103	5.83	-50.9107
1738	11.375	000100011011	MOEBIUS	-0.1427	-1.31	-69.2792
2489	12.2926	001101100111	MOEBIUS	-0.1766	-1.00	-70.2633
266	13.8368	000110011011	HUECKEL	0.0777	2.07	-60.7019
6513	14.5144	000100010001	MOEBIUS	-0.3754	-0.21	-68.6185
2960	15.2592	000100110011	MOEBIUS	0.1938	-5.90	-76.8169
2880	16.282	000000010001	HUECKEL	-0.1947	-1.92	-72.8469
2896	17.4955	000010011001	HUECKEL	-0.1557	-2.13	-68.9712
2916	18.1087	000101000101	HUECKEL	-0.3606	-0.48	-69.4919
218	18.8539	000011001111	HUECKEL	0.0851	3.68	-55.643
2261	20.4208	000010001101	HUECKEL	0.0206	-1.84	-69.7781
1537	21.93	000111000111	HUECKEL	-0.044	-0.85	-65.7178
3103	24.3219	001010011011	HUECKEL	-0.1761	0.78	-66.899
2037	25.3686	001111001111	HUECKEL	-0.0232	3.31	-59.6327
1809	26.339	001101100111	MOEBIUS	-0.0299	-0.60	-69.3246
3158	26.3495	000101001111	HUECKEL	-0.0764	-2.30	-68.3437
2535	27.2768	001011001111	MOEBIUS	-0.0036	-1.90	-70.7391
2059	30.7167	001101111111	MOEBIUS	0.1901	-7.25	-76.7872
3184	31.6342	001011011011	MOEBIUS	0.1076	-4.95	-75.6223
4180	33.8576	011011011011	HUECKEL	-0.2119	3.46	-65.3169
1803	34.3535	001011001011	HUECKEL	-0.2963	0.16	-68.5112
4488	34.9774	011011011011	HUECKEL	-0.079	2.46	-64.6917
5242	38.3901	011011011011	HUECKEL	-0.3004	0.39	-67.1756
1580	38.8165	000101111011	MOEBIUS	0.3239	-6.55	-74.7085

\*: von SEQLOG2XYZ ausgegebene NICS Werte entsprechen der magnetischen Abschirmung aus den GAUSSIAN Ausgabedateien, also vorzeichenverkehrt zur Literaturkonvention.<sup>[73]</sup> In Tabelle 81 wird der vorzeichenkorrigierte Wert angegeben.

### 3.5.7 FISE-Rechnungen

Wie bereits beschrieben ist die ISE<sub>II</sub>-Methode für Isomere in Möbiustopologie nicht durchführbar. Es wird die FISE-Methode nach Kapitel 1.4.3 verwendet.

OPT\_FINDEN:

```
12/mmp/$
xyz2cane -s sel/c12ringtors.sel -i
c12_mmp_opt_b3lyp_6-31Gd.cane.xyz -c -m 8

12/mmp/b3lyp_6-31Gd/opt_finden/$
cp ../../c12_mmp_opt_b3lyp_6-31Gd.cane_inden.xyz
c12_mmp_opt_finden_b3lyp_6-31Gd.xyz

12/mmp/b3lyp_6-31Gd/opt_finden/$
cp ../../c12_mmp_opt_b3lyp_6-31Gd.cane_ise.fixed
c12_mmp_opt_finden_b3lyp_6-31Gd.fixed

12/mmp/b3lyp_6-31Gd/opt_finden/$
xyz2com -k 'b3lyp/6-31G* opt=modredundant
geom=(nodist, noangle) nopop' -c 0 -m 1 -g g03 -f2
c12_mmp_opt_finden_b3lyp_6-31Gd.fixed -s
c12_mmp_opt_finden_b3lyp_6-31Gd.xyz

12/mmp/b3lyp_6-31Gd/opt_finden/$
nohup sh c12_mmp_opt_finden_b3lyp_6-31Gd.sh &

12/mmp/b3lyp_6-31Gd/opt_finden/$
seqlog2xyz -s .log -a 1 -z 10000
c12_mmp_opt_finden_b3lyp_6-31Gd. >
c12_mmp_opt_finden_b3lyp_6-31Gd.seq.xyz

12/mmp/ $
xyz2cane -s sel/c12ringtors.sel -i b3lyp_6-
31Gd/opt_finden/c12_mmp_opt_finden_b3lyp_6-
31Gd.seq.xyz -o c12_mmp_opt_finden_b3lyp_6-
31Gd.cane.xyz -v -c > c12_mmp_opt_finden_b3lyp_6-
31Gd.cane.csv
```

Mit obigen Kommandozeilen werden die GAUSSIAN *input files* für eine Optimierung mit eingefrorenem Ring und optimierbarer Indeneinheit sowie Wasserstoffatomen erzeugt und gestartet. Analog für die Isoindenderivatisierung:

OPT\_FISOINDEN:

```
12/mmp/b3lyp_6-31Gd/opt_fisoinden/$
cp ../../c12_mmp_opt_b3lyp_6-31Gd.cane_isoinden.xyz
c12_mmp_opt_fisoinden_b3lyp_6-31Gd.xyz
```

```

12/mmp/b3lyp_6-31Gd/opt_fisoinden/$
cp ../../c12_mmp_opt_b3lyp_6-31Gd.cane_ise.fixed
c12_mmp_opt_fisoinden_b3lyp_6-31Gd.fixed

12/mmp/b3lyp_6-31Gd/opt_fisoinden/$
xyz2com -k 'b3lyp/6-31G* opt=modredundant
geom=(nodist, noangle) nopop' -c 0 -m 1 -g g03 -f2
c12_mmp_opt_fisoinden_b3lyp_6-31Gd.fixed -s
c12_mmp_opt_fisoinden_b3lyp_6-31Gd.xyz

12/mmp/b3lyp_6-31Gd/opt_fisoinden/$
nohup sh c12_mmp_opt_fisoinden_b3lyp_6-31Gd.sh &

12/mmp/b3lyp_6-31Gd/opt_fisoinden/$
seqlog2xyz -s .log -a 1 -z 10000
c12_mmp_opt_fisoinden_b3lyp_6-31Gd. >
c12_mmp_opt_fisoinden_b3lyp_6-31Gd.seq.xyz

12/mmp/ $ xyz2cane -s sel/c12ringtors.sel -i b3lyp_6-
31Gd/opt_fisoinden/c12_mmp_opt_fisoinden_b3lyp_6-
31Gd.seq.xyz -o c12_mmp_opt_fisoinden_b3lyp_6-
31Gd.cane.xyz -v -c > c12_mmp_opt_fisoinden_b3lyp_6-
31Gd.cane.csv

```

Es folgt die Zusammenführung der Ergebnisse ins xyz-Format und Auswertung mittels Xyz2cane.

```

12/mmp/$
seqlog2xyz b3lyp_6-31Gd/opt/c12_mmp_opt_b3lyp_6-31Gd.
-j b3lyp_6-
31Gd/opt_finden/c12_mmp_opt_finden_b3lyp_6-31Gd. -k
b3lyp_6-
31Gd/opt_fisoinden/c12_mmp_opt_fisoinden_b3lyp_6-
31Gd. -m b3lyp_6-31Gd/mags/c12_mmp_mags_b3lyp_6-31Gd.
-n b3lyp_6-31Gd/nics/c12_mmp_nics_b3lyp_6-31Gd. >
c12_mmp_fiseopt_b3lyp_6-31Gd.seq.xyz

12/mmp/$
xyz2cane -s sel/c12ringtors.sel -i
c12_mmp_fiseopt_b3lyp_6-31Gd.seq.xyz -o
c12_mmp_fiseopt_b3lyp_6-31Gd.cane.xyz -v -c >
c12_mmp_fiseopt_b3lyp_6-31Gd.cane.csv

```

Neu hinzugekommen zu den bisherigen Tabellenwerten sind jetzt die *fixed indene energy*, die *fixed isoindene energy* sowie deren Differenz, die FISE-Werte, wiedergegeben in Tabelle 82.

### 3.5.8 FISE-NICS und -MAGS

Auch für die *fixed indene* Strukturen werden die NICS- und MAGS-Werte berechnet.

## NICS – FINDEN:

```
12/mmp/ $ xyz2cane -s sel/cl2ringtors.sel -i
          c12_mmp_opt_finden_b3lyp_6-31Gd.cane.xyz -c -n
          c12_mmp_nics_finden_b3lyp_6-31Gd.xyz

12/mmp/b3lyp_6-31Gd/nics_finden $
          cp ../../c12_mmp_nics_finden_b3lyp_6-31Gd.xyz .

12/mmp/b3lyp_6-31Gd/nics_finden $
          xyz2com -k 'b3lyp/6-31G* NMR=GIAO' -c 0 -m 1 -g g03 -
          s c12_mmp_nics_finden_b3lyp_6-31Gd.xyz

12/mmp/b3lyp_6-31Gd/nics_finden $
          nohup sh c12 c12_mmp_nics_finden_b3lyp_6-31Gd.sh &
```

## NICS – FISOINDEN:

```
12/mmp/ $ xyz2cane -s sel/cl2ringtors.sel -i
          c12_mmp_opt_fisoinden_b3lyp_6-31Gd.cane.xyz -c -n
          c12_mmp_nics_fisoinden_b3lyp_6-31Gd.xyz

12/mmp/b3lyp_6-31Gd/nics_fisoinden $
          cp ../../c12_mmp_nics_fisoinden_b3lyp_6-31Gd.xyz .

12/mmp/b3lyp_6-31Gd/nics_fisoinden $
          xyz2com -k 'b3lyp/6-31G* NMR=GIAO' -c 0 -m 1 -g g03 -
          s c12_mmp_nics_fisoinden_b3lyp_6-31Gd.xyz

12/mmp/b3lyp_6-31Gd/nics_fisoinden $
          nohup sh c12 c12_mmp_nics_fisoinden_b3lyp_6-31Gd.sh &
```

## MAGS – FINDEN:

```
12/mmp/b3lyp_6-31Gd/mags_finden $
          cp ../../c12_mmp_opt_finden_b3lyp_6-31Gd.cane.xyz
          c12_mmp_mags_finden_b3lyp_6-31Gd.xyz

12/mmp/b3lyp_6-31Gd/mags $
          xyz2com -k 'hf/6-31G* NMR=CSGT' -c 0 -m 1 -g g03 -s
          c12_mmp_mags_finden_b3lyp_6-31Gd.xyz

12/mmp/b3lyp_6-31Gd/mags $
          nohup sh c12_mmp_mags_finden_b3lyp_6-31Gd.sh &
```

## MAGS – FISOINDEN:

```
12/mmp/b3lyp_6-31Gd/mags_fisoinden $
          cp ../../c12_mmp_opt_fisoinden_b3lyp_6-31Gd.cane.xyz
          c12_mmp_mags_fisoinden_b3lyp_6-31Gd.xyz

12/mmp/b3lyp_6-31Gd/mags $
          xyz2com -k 'hf/6-31G* NMR=CSGT' -c 0 -m 1 -g g03 -s
          c12_mmp_mags_fisoinden_b3lyp_6-31Gd.xyz
```

```
12/mmp/b3lyp_6-31Gd/mags $
  nohup sh c12_mmp_mags_fisoinden_b3lyp_6-31Gd.sh &
```

Die Auswertung erfolgt in separaten csv-Tabellen für die *fixed indene* und die *fixed isoindene* Derivatisierung. Die Bildung der Differenzen der NICS- und MAGS-Werte erfolgt anschließend manuell nach Gleichung (1.31) und (1.32).

FINDEN:

```
12/mmp/$
  seqlog2xyz b3lyp_6-31Gd/opt/c12_mmp_opt_b3lyp_6-31Gd.
  -m b3lyp_6-
  31Gd/mags_finden/c12_mmp_mags_finden_b3lyp_6-31Gd. -n
  b3lyp_6-31Gd/nics_finden/c12_mmp_nics_finden_b3lyp_6-
  31Gd. > c12_mmp_nicsmags_finden_b3lyp_6-31Gd.seq.xyz

12/mmp/ $
  xyz2cane -s sel/c12ringtors.sel -i
  c12_mmp_nicsmags_finden_b3lyp_6-31Gd.seq.xyz -o
  c12_mmp_nicsmags_finden_b3lyp_6-31Gd.cane.xyz -v -c >
  c12_mmp_nicsmags_finden_b3lyp_6-31Gd.cane.csv
```

FISOINDEN:

```
12/mmp/$
  seqlog2xyz b3lyp_6-31Gd/opt/c12_mmp_opt_b3lyp_6-31Gd.
  -m b3lyp_6-
  31Gd/mags_fisoinden/c12_mmp_mags_fisoinden_b3lyp_6-
  31Gd. -n b3lyp_6-
  31Gd/nics_fisoinden/c12_mmp_nics_fisoinden_b3lyp_6-
  31Gd. > c12_mmp_nicsmags_fisoinden_b3lyp_6-
  31Gd.seq.xyz

12/mmp/ $
  xyz2cane -s sel/c12ringtors.sel -i
  c12_mmp_nicsmags_fisoinden_b3lyp_6-31Gd.seq.xyz -o
  c12_mmp_nicsmags_fisoinden_b3lyp_6-31Gd.cane.xyz -v -
  c > c12_mmp_nicsmags_fisoinden_b3lyp_6-31Gd.cane.csv
```

Die wichtigsten Daten der vollständigen KONAN Analyse ist in Tabelle 82 wiedergegeben.



Tabelle 82: [12]Annulen – wichtigste Aromatizitätskriterien (DFT-opt)

lognr	E <sub>rel</sub>	CANE string	T**	T <sub>max</sub>	R <sub>av</sub>	A <sub>J</sub>	HOMA	NICS*	MAGS	FISE	FISE-NICS	FISE-MAGS
146	0.00	000010100011	H	62.48	1.4084	0.663	0.127	-6.21	-72.29	3.34	-1.56	-4.96
61	1.63	000001000001	H	40.93	1.4068	0.7156	0.264	-6.95	-75.46	3.50	-1.03	-3.16
37	1.86	001100110011	H	59.62	1.4109	0.6405	0.045	3.39	-60.96	0.66	4.02	6.21
639	5.37	000111000111	H	80.53	1.4097	0.5566	-0.131	-4.53	-69.60	2.14	-0.11	-2.01
13	5.43	000001100011	H	72.86	1.4086	0.6029	-0.012	-2.89	-67.97	2.66	0.71	0.23
674	5.48	000111000111	H	50.55	1.4099	0.5863	-0.066	-4.18	-68.38	0.62	0.55	0.21
3327	6.16	000001100011	H	79.64	1.4088	0.6125	0.008	-2.53	-72.37	3.36	0.47	-3.07
421	6.25	000001000011	M	53.62	1.4052	0.9009	0.700	-14.69	-95.69	10.99	-10.38	-28.03
1	6.68	000011000011	H	37.35	1.4084	0.6884	0.185	4.17	-48.16	-1.35	8.16	15.10
194	6.73	000010000111	H	59.14	1.4089	0.6036	-0.014	-2.02	-65.33	2.01	1.26	0.69
1454	7.37	000110010011	M	81.96	1.4116	0.6017	-0.053	0.04	-65.32	0.17	0.82	-0.27
1212	8.90	000000100011	M	74.75	1.4086	0.6074	-0.002	-2.32	-70.31	4.66	-1.52	-4.19
121	9.78	000110001111	H	52.89	1.4097	0.6574	0.099	-0.14	-58.40	1.14	6.11	10.03
358	10.41	000100110011	M	79.55	1.4109	0.6051	-0.035	-1.66	-71.46	3.00	-0.32	-2.93
568	11.01	000110010011	M	69.77	1.4115	0.6139	-0.024	0.23	-65.79	0.28	0.49	-1.16
1007	11.13	000110011011	H	55.70	1.4112	0.6709	0.110	5.83	-50.91	0.47	10.11	18.22
1738	11.38	000100011011	M	78.80	1.4119	0.5642	-0.143	-1.31	-69.28	-0.46	-0.35	-2.30
2489	12.29	001101100111	M	80.47	1.4126	0.5536	-0.177	-1.00	-70.26	-0.96	-0.54	-3.00
266	13.84	000110011011	H	62.02	1.4124	0.6632	0.078	2.07	-60.70	2.63	4.01	6.60
6513	14.51	000100010001	M	69.92	1.4117	0.4611	-0.375	-0.21	-68.62	1.20	0.00	-1.95
2960	15.26	000100110011	M	56.40	1.4099	0.7004	0.194	-5.90	-76.82	2.39	-3.96	-11.12
2880	16.28	000000010001	H	77.44	1.4101	0.5306	-0.195	-1.92	-72.85	2.10	-0.83	-3.55
2896	17.50	000010011001	H	84.90	1.4127	0.5634	-0.156	-2.13	-68.97	1.89	0.19	-0.61
2916	18.11	000101000101	H	70.14	1.4125	0.4722	-0.361	-0.48	-69.49	-1.69	0.89	-0.61
218	18.85	000011001111	H	54.69	1.4111	0.6592	0.085	3.68	-55.64	0.69	6.18	9.08
2261	20.42	000010001101	H	70.09	1.4117	0.6346	0.021	-1.84	-69.78	1.96	0.76	-0.24
1537	21.93	000111000111	H	53.90	1.4113	0.6034	-0.044	-0.85	-65.72	-0.21	1.47	0.46
3103	24.32	001010011011	H	55.48	1.4147	0.5672	-0.176	0.78	-66.90	1.01	1.57	1.03
2037	25.37	001111001111	H	47.49	1.4144	0.6319	-0.023	3.31	-59.63	-1.68	4.74	6.03
1809	26.34	001101100111	M	69.73	1.4147	0.6312	-0.030	-0.60	-69.32	-0.56	-0.02	-0.97
3158	26.35	000101001111	H	66.43	1.4133	0.6016	-0.076	-2.30	-68.34	0.82	1.29	-0.50
2535	27.28	001011001111	M	80.50	1.4136	0.6352	-0.004	-1.90	-70.74	1.77	-0.85	-4.08
2059	30.72	001101111111	M	87.55	1.4116	0.7082	0.190	-7.25	-76.79	2.22	-4.11	-10.48
3184	31.63	001011011011	M	68.87	1.4143	0.6885	0.108	-4.95	-75.62	2.61	-2.93	-7.57
4180	33.86	011011011011	H	45.13	1.4186	0.5791	-0.212	3.46	-65.32	-3.32	2.39	3.27
1803	34.35	001011001011	H	79.81	1.4162	0.525	-0.296	0.16	-68.51	-3.08	0.41	-0.82
4488	34.98	011011011011	H	54.07	1.4185	0.6356	-0.079	2.46	-64.69	-0.08	2.66	4.41
5242	38.39	011011011011	H	72.33	1.4162	0.523	-0.300	0.39	-67.18	-4.02	0.86	0.31
1580	38.82	000101111011	M	89.51	1.4109	0.763	0.324	-6.55	-74.71	3.01	-4.20	-9.85

\*siehe Tabelle 81, \*\*T=Topology, H=Hückel, M=Möbius, Energien in kcal/mol

### 3.5.9 Korrelationsgraphen

Ein vollständiger Satz an Auftragungen der Werte für E<sub>rel</sub>, FISE, T<sub>max</sub>, R<sub>av</sub> und der Aromatizitätskriterien HOMA, NICS, MAGS, FISE-NICS, FISE-MAGS ist im Anhang unter dem Namen /data/annulen/12/anhang-c12-graphen.doc enthalten.

Katritzky *et al.* vertreten durch PCA (*principal component analysis*, siehe Kapitel 1.3.4) gestützt die Auffassung,<sup>[89]</sup> dass es zwei orthogonale Arten von

Aromatizität gibt. Zum einen ist dies die „klassische Aromatizität“, hauptsächlich charakterisiert durch geometrische und energetische Kriterien, zum anderen die „magnetische Aromatizität“, charakterisiert durch magnetische Kriterien. Im Folgenden wird die Korrelation der einzelnen Klassen durch Auftragung je zweier Kriterien, getrennt nach den Hückel- und Möbiusstrukturen, überprüft.

### 3.5.9.1 Strukturelle Kriterien

Als strukturelle Aromatizitätskriterien zeigen die verwendeten Kriterien HOMA und der Julg-Index  $A_j$  eine lineare Abhängigkeit.

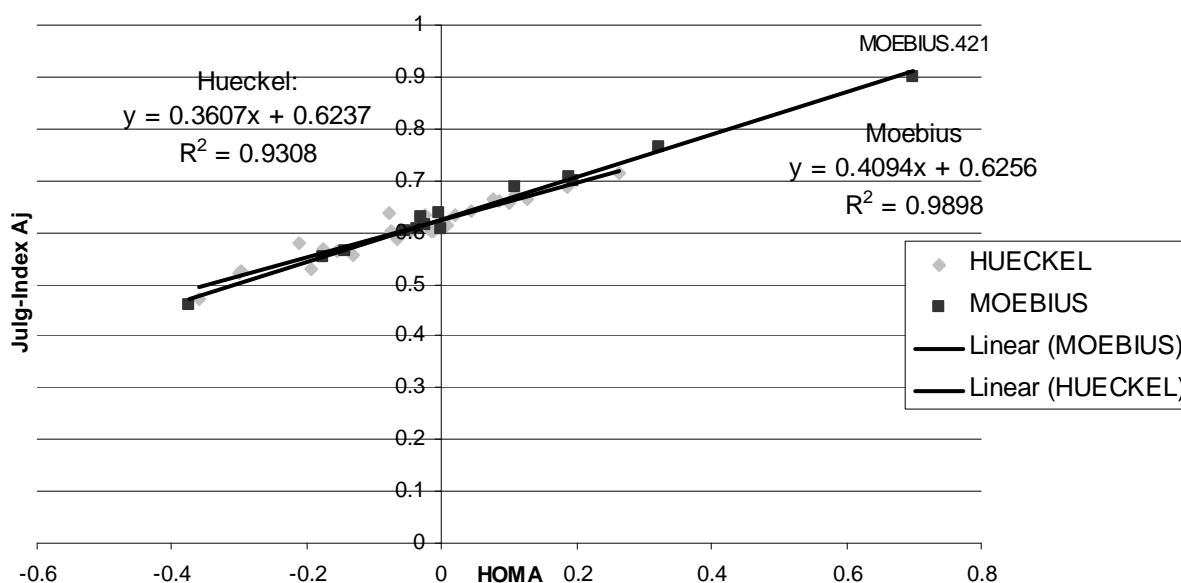


Abbildung 178: [12]Annulen, Julg-Index und HOMA

Diese Beziehung ist praktisch unabhängig von der Topologie der Strukturen. Bis auf eine Möbiusstruktur mit HOMA von 0.7 (lognr 421, die stabilste Struktur mit Möbius-Topologie) liegen alle Datenpunkte im schwach bis nicht aromatischen Bereich. Zum Vergleich sei auf Tabelle 8 verwiesen.

Damit ist das hauptsächliche Problem der Aromatizitätsuntersuchung von [4n]Annulenen auch schon benannt. Die Effekte sind relativ schwach im Vergleich zu klassischen Hückelaromaten und werden durch andere Einflussgrößen wie sterische Hinderung und Deplanarisierung überlagert. Abbildung 179 belegt, dass anhand der strukturellen Kriterien Aromatizität in [4n]Annulenen nicht nachgewiesen werden kann.<sup>7</sup>

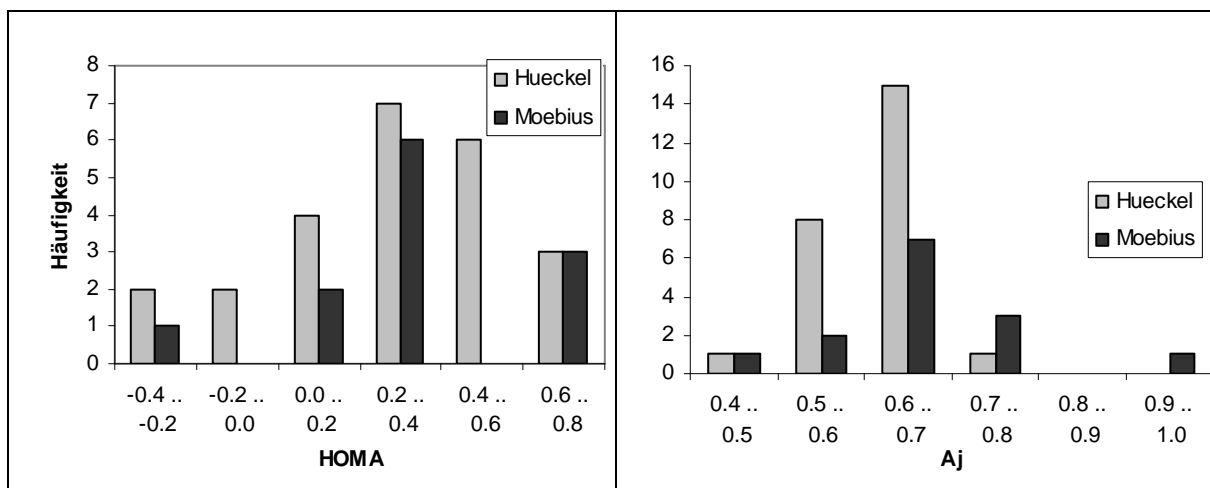


Abbildung 179: [12]Annulene, Histogramme für HOMA links und  $A_j$  rechts

Auch die Kombination energetischer Kriterien - wie  $E_{rel}$  oder FISE - und struktureller Kriterien ist nicht geeignet, eventuelle aromatische Effekte der [4n]Annulene zu charakterisieren, siehe Abbildung 180.

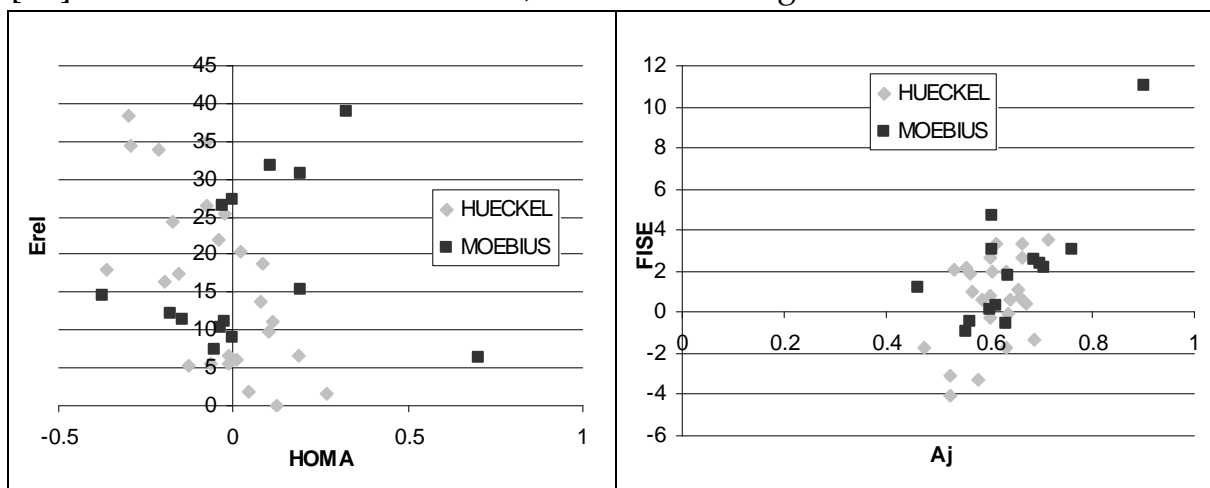


Abbildung 180: [12]Annulene,  $E_{rel}$  über HOMA links, FISE über  $A_j$  rechts

Eine Ausnahme stellt wiederum die klar aromatische Möbiusstruktur 421 dar.

### 3.5.9.2 Energetische Kriterien

Die Verteilung der B3LYP/6-31G\*-Energieniveaus für Möbius- und Hückelannulene sind der Abbildung 181 zu entnehmen. Während die absoluten Energien durch andere Einflüsse als die Topologie dominiert werden, zeigen die FISE-Werte im Einzelfall für die Möbiusstruktur 421 eine klare aromatische Stabilisierung mit einem Wert von +12 kcal/mol.

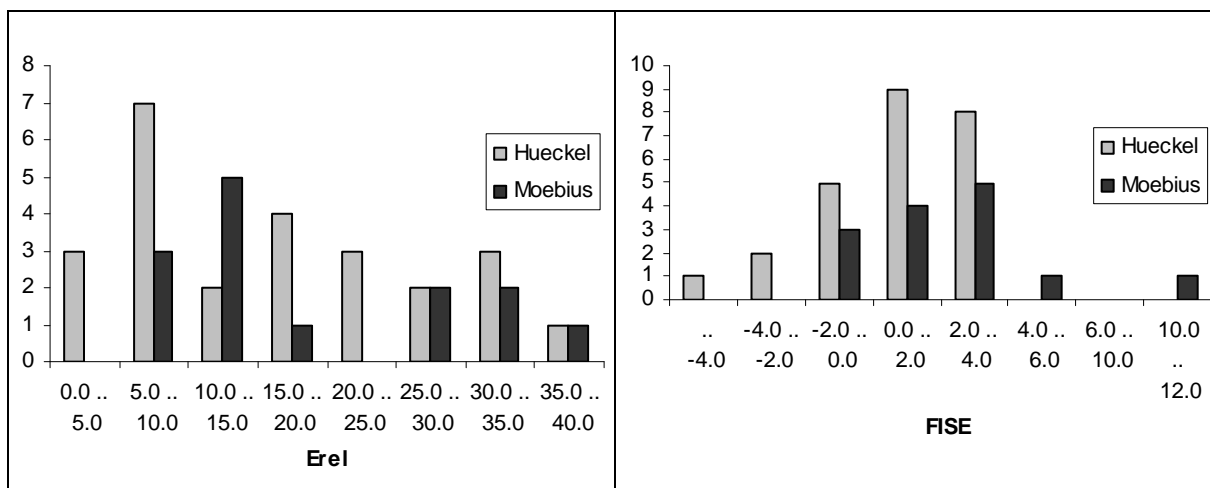


Abbildung 181: [12]Annulen, Histogramme  $E_{rel}$  links und FISE rechts

Bei den Hückelstrukturen gibt es einige Isomere mit negativer, sprich antiaromatischer Stabilisierungsenergie. Insgesamt befinden sich jedoch der Großteil der Datenpunkte in einem Bereich um die 2-4 kcal/mol, was im Bereich der linearen Polyen-Konjugationsstabilisierungsenergie (CSE) liegt (siehe Abbildung 15 und Tabelle 7) und eine eindeutige Zuordnung als aromatisch bzw. antiaromatisch nicht ermöglicht. Die CSE der Möbiusisomere ist im Durchschnitt größer, was zumindest für einen schwachen aromatischen Anteil an der CSE spricht.

Eine Auftragung der Energiewerte über strukturelle Indikatoren liefert denn auch ein relativ wenig aussagekräftiges Bild, siehe Abbildung 182.

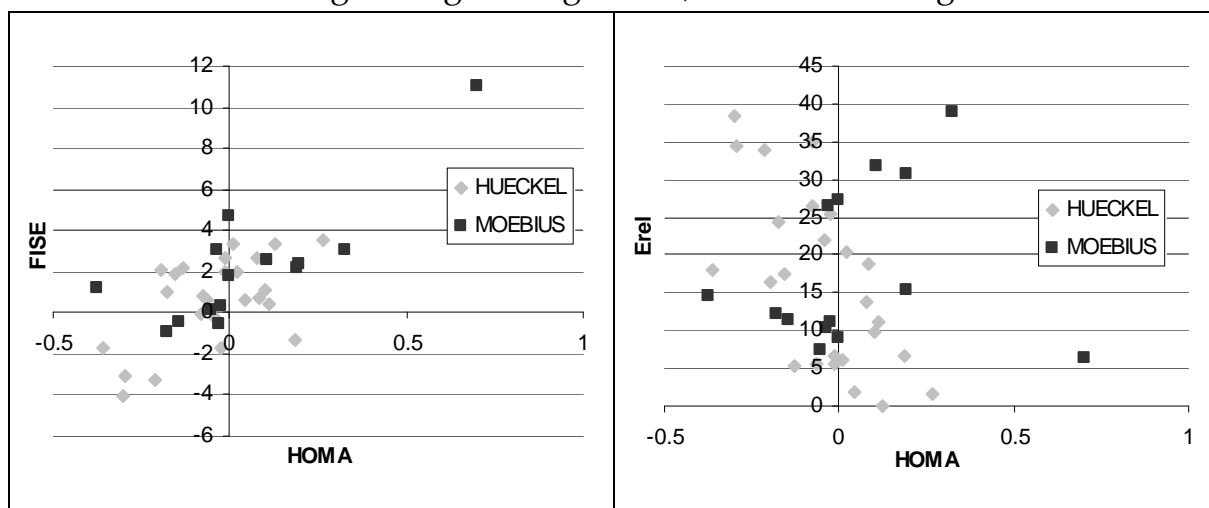
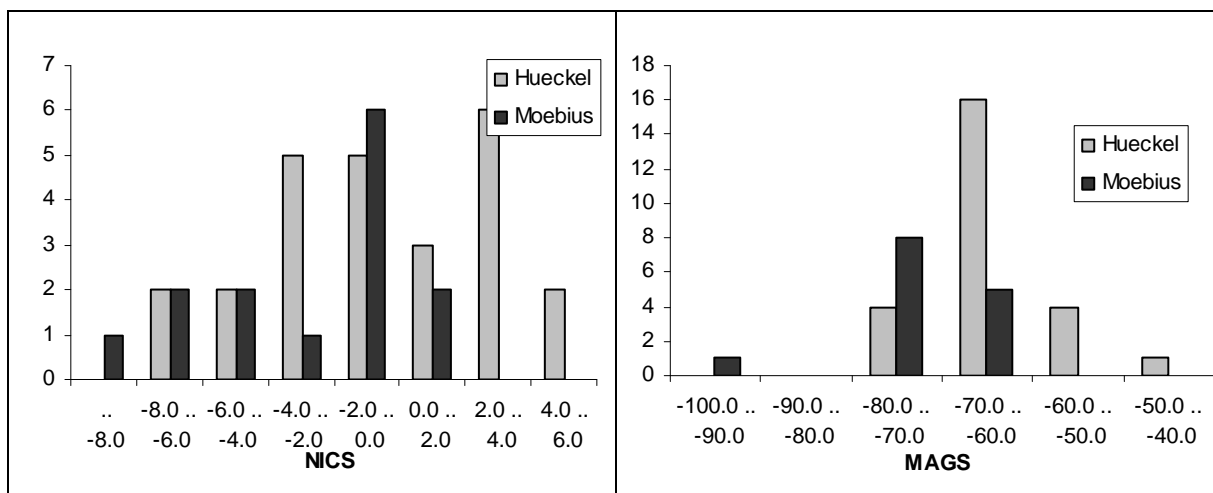


Abbildung 182: [12]Annulen, FISE und Erel über HOMA

### 3.5.9.3 Magnetische Kriterien

Als magnetische Aromatizitätskriterien werden NICS und MAGS, sowie die neu eingeführten FISENICS und FISEMAGS, verwendet. Dabei zeigt sich der eingeschränkte Nutzen der bekannten magnetischen Aromatizitätskriterien

zum Vergleich unterschiedlicher Isomere des gleichen Moleküls in Abbildung 183.



**Abbildung 183:** [12]Annulen, Histogramme NICS (links) und MAGS (rechts)

Interessant ist wiederum ein Vergleich mit den Werten des Benzols als „Musteraromat“.

$$\text{NICS}(\text{C}_6\text{H}_6, \text{B3LYP}/6\text{-}31\text{G}^*) = -9.68 \text{ ppm}$$

$$\text{MAGS}(\text{C}_6\text{H}_6, \text{HF}/6\text{-}31\text{G}^*) = -49.06 \text{ cgs-ppm}$$

Der NICS-Wert des MOEBIUS.421 Isomers liegt mit  $-14.59 \text{ ppm}$  noch niedriger als der von Benzol, was die Zuverlässigkeit von NICS-Werten als Aromatizitätskriterium in Frage stellt. Eine mögliche Erklärung liegt in der ausgeprägten 3D-Struktur der meisten Isomere, wodurch das NICS-Zentrum in das lokale Feld einzelner Atome geraten kann. Dies ist ein systematisches Problem der NICS-Methode für alle nicht planaren Moleküle.

Die NICS-Verteilung der Hückel- und der Möbiusisomere entspricht näherungsweise zwei Gaußfunktionen, das Maximum der „aromatischen“ Möbiusisomere liegt wie erwartet bei tieferen Werten als das der Hückelisomere. Da die Maxima zu dicht beieinander liegen, sind die NICS-Werte zur Unterscheidung des aromatischen und antiaromatischen Charakters der Isomere jedoch generell nicht besonders geeignet. Die MAGS-Werte zeigen das gleiche Verhalten, allerdings sind die Maxima besser getrennt.

Außerdem teilt die magnetische Suszeptibilität als Eigenschaft des Gesamtsystems die lokale Störungsproblematik der NICS-Werte nicht. Werte von unterschiedlichen Molekülen lassen sich jedoch nicht miteinander vergleichen, siehe die angegebenen Werte des Benzols.

Wie erwartet sind die neu eingeführten FISENICS und FISEMAGS einfacher zu deuten, siehe Abbildung 184.

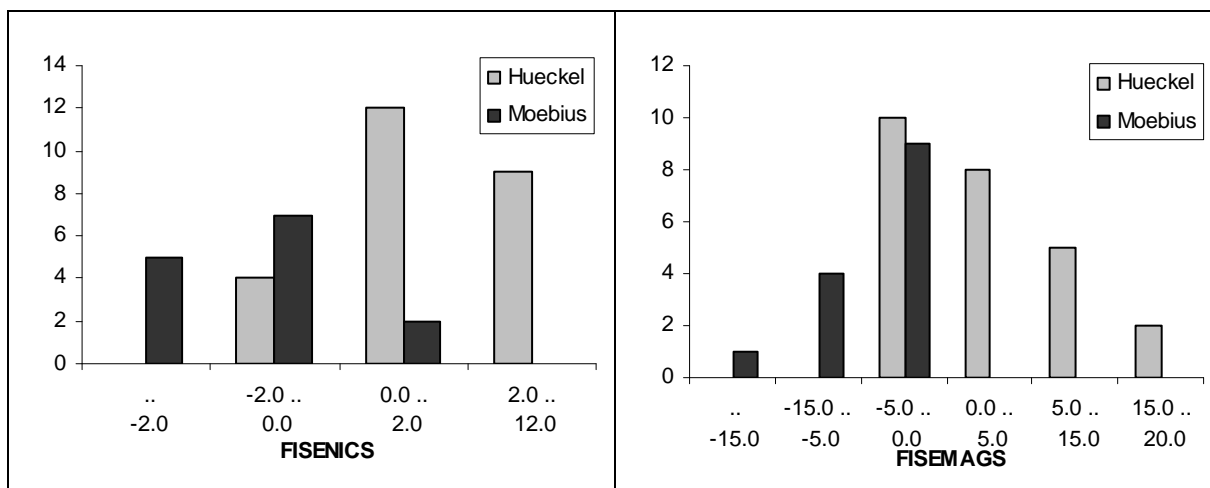


Abbildung 184: [12]Annulen, Histogramme FISENICS (links) und FISEMAGS (rechts)

Negative Werte stehen hier für eine aromatische Stabilisierung (im ursprünglichen, nicht derivatisierten, Molekül) durch Konjugation über die FISE-derivatisierte Doppelbindung, positive Werte für eine antiaromatische Destabilisierung.

Eine Auftragung der NICS- und MAGS-Werte über einen strukturellen Aromatizitätsindikator zeigt den gegenläufigen Trend dieser Indikatoriklassen (siehe Abbildung 185).

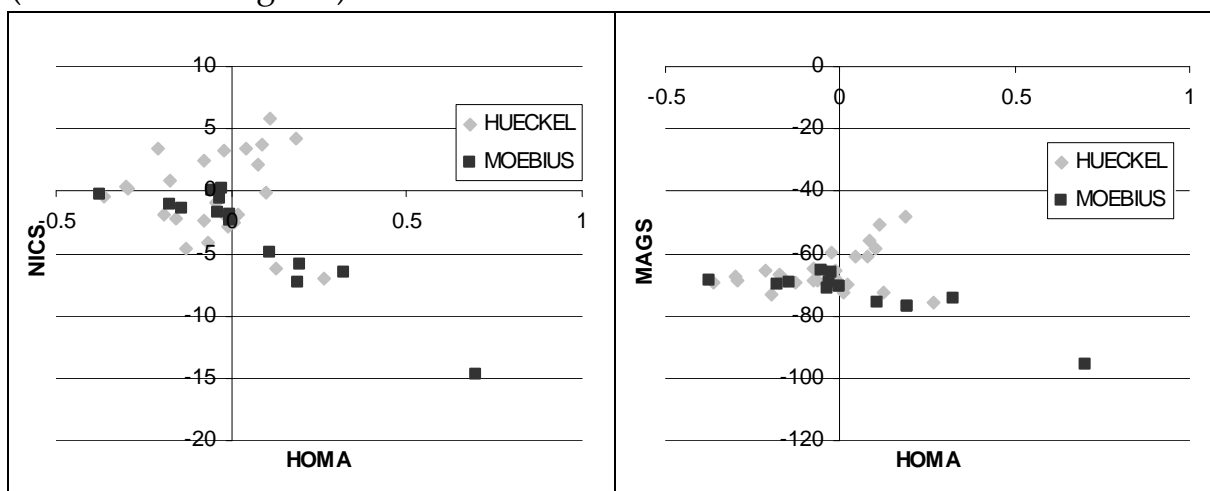
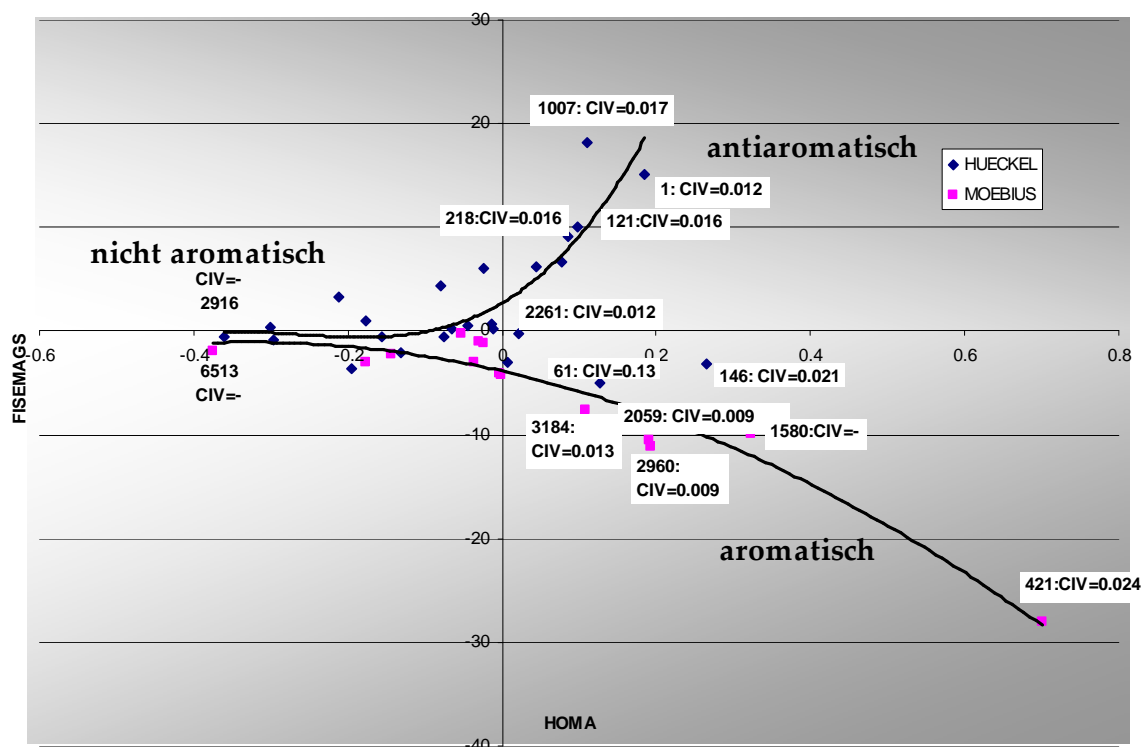


Abbildung 185: [12]Annulen, Auftragung der NICS- und MAGS-Werte über die HOMA Werte

Problematisch ist die relativ große Streuung bei den NICS-Werten, was zu einer schlechten Trennung der aromatischen Möbiusisomeren von den antiaromatischen Hückelstrukturen führt. Daraus resultierend können die meisten Strukturen nicht zufriedenstellend klassifiziert werden. Bei den MAGS-Werten ist die Bereichsvariabilität das größte Problem, da nur für ein größeres Ensemble von Wertepaaren einzelne Werte klassifiziert werden können.

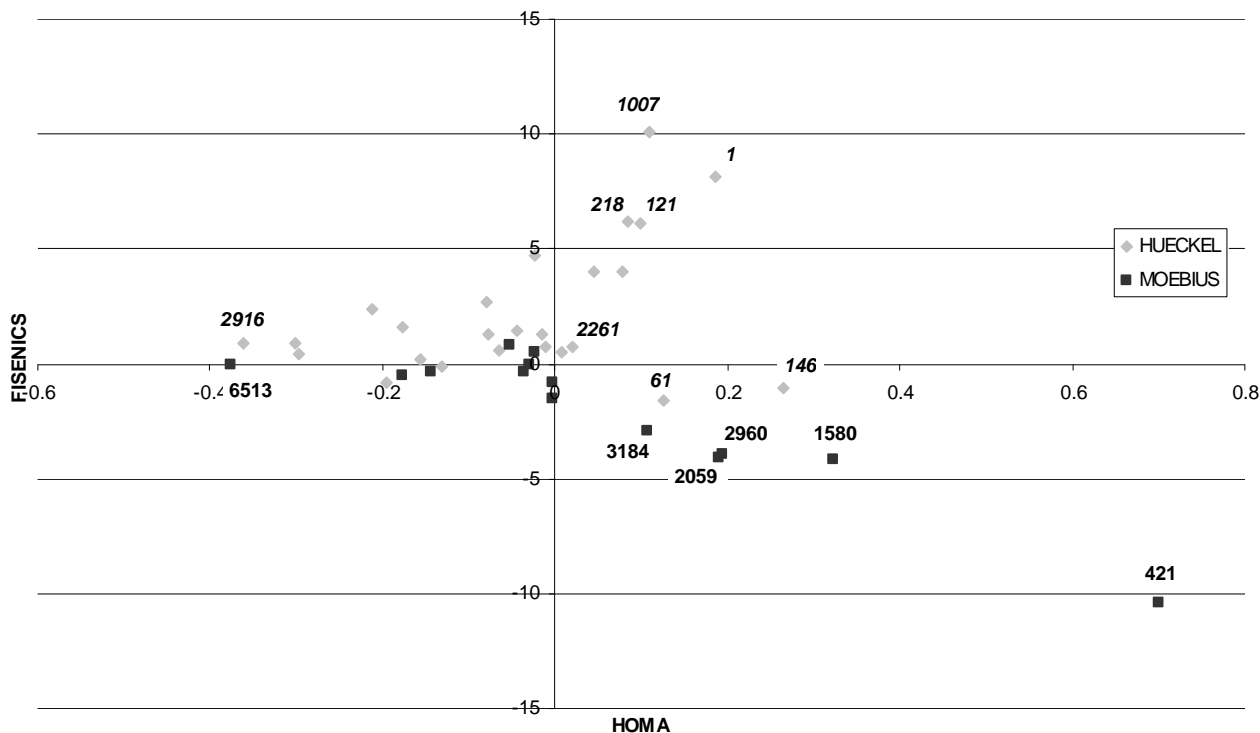
Eine Auftragung der FISEMAGS-Werte über die HOMA Werte, siehe Abbildung 186, ist erwartungsgemäß besser geeignet, Strukturen in nichtaro-

matisch, antiaromatisch und aromatisch einzuordnen. Dabei entziehen sich die nichtaromatischen Strukturen durch Bindungslängenalternanz und/oder starke Abweichungen einzelner Doppelbindungen aus der Planarität (Torsionswinkel der Doppelbindung bei ca. 90°) der „aromatischen Überlappung“.



**Abbildung 186:** [12]Annulen, Auftragung der FISEMAGS-Werte über die HOMA-Werte sowie die CIV-Werte ausgesuchter Strukturen

Als letzte Grafik dieses Kapitels verdeutlicht die Auftragung der FISENICS – Werte, dass diese prinzipiell in Kombination mit einem strukturellen Kriterium ebenfalls zur Klassifizierung geeignet sind.



**Abbildung 187:**  $[12]$ Annulene, Auftragung der FISENICS-Werte über die HOMA-Werte mit Zuordnung ausgewählter Strukturen

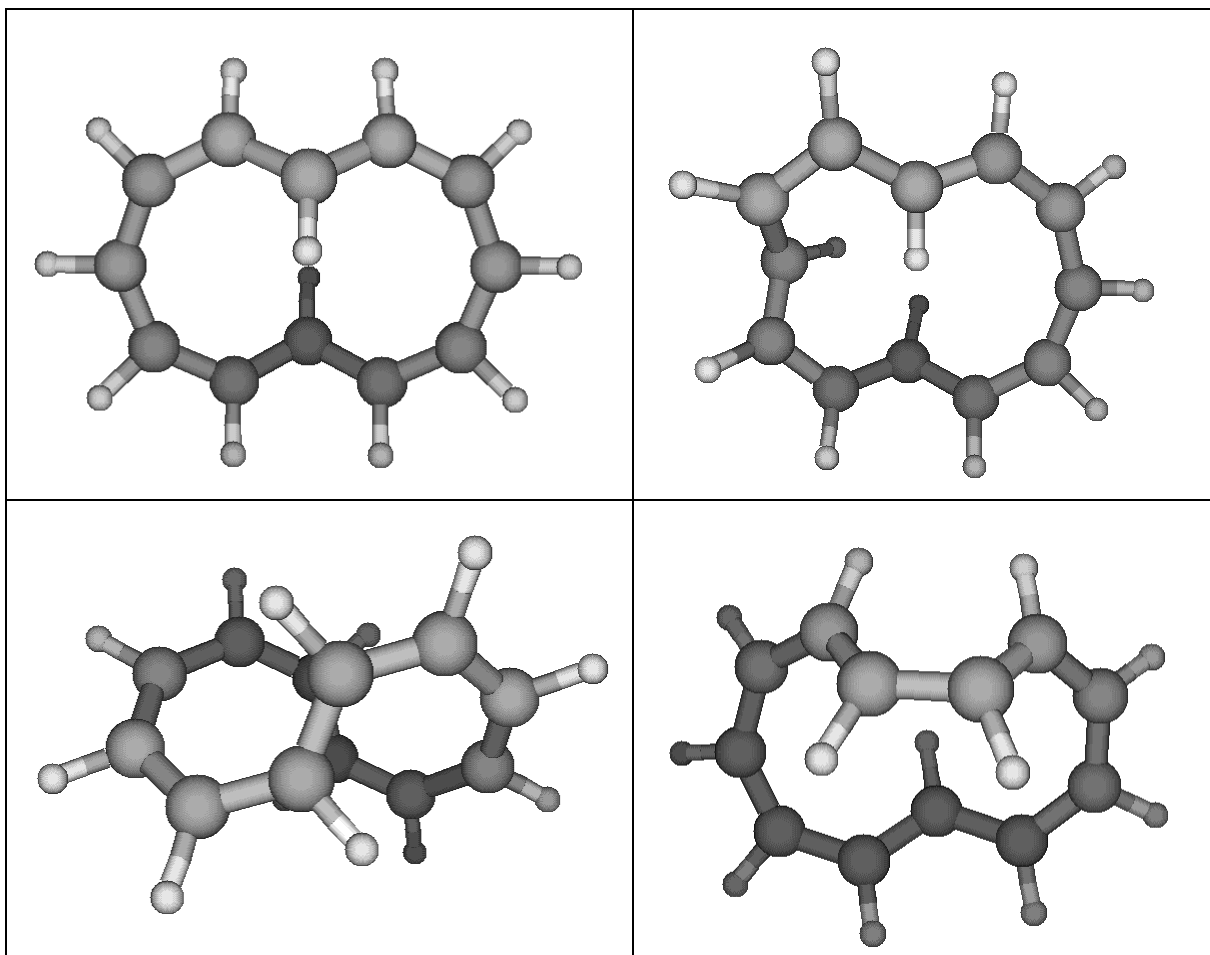
Offensichtlich existiert Möbiusaromatizität bzw. Hückelantiaromatizität für  $[4n]$ Annulene durchaus. Dies wurde bisher zwar postuliert,<sup>[64]</sup> konnte aber noch nicht schlüssig experimentell oder rechnerisch erhärtet werden.

Diskussionsbedarf besteht für die Hückelisomere 61 und 146, welche sich augenscheinlich als Ausreißer in den aromatischen Bereich „verirrt“ haben. Eine genauere Betrachtung der elektronischen Situation in diesen Strukturen soll mit Hilfe der AICD-Funktion diese scheinbaren Widersprüche im folgenden Kapitel aufklären.

### 3.5.10 AICD-Plots

Die unerwartete Aromatizität einiger Hückelstrukturen wurde bereits im letzten Abschnitt thematisiert. Zum Vergleich sind die beiden antiaromatischen Hückelstrukturen gegenüber den beiden aromatischsten in Abbildung 188 angezeigt.

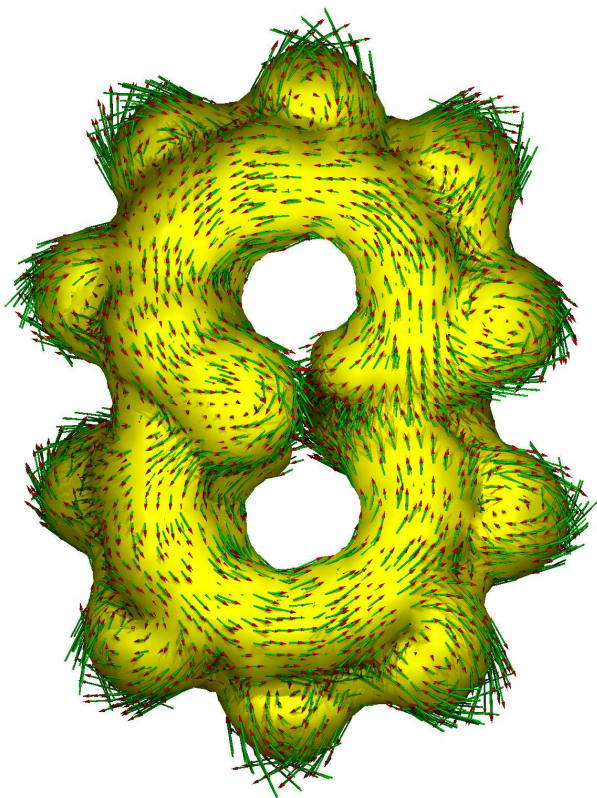




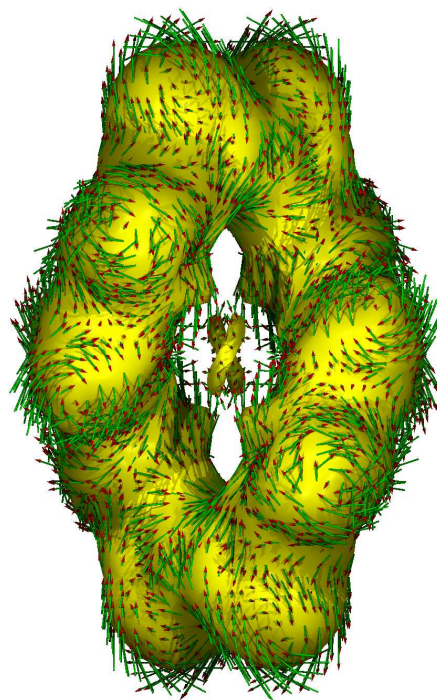
**Abbildung 188:** dreidimensionale Strukturen:  
antiaromatische Hückelisomere 1 und 1007 (oben)  
aromatische Hückelisomere 61 und 146 (unten)

Die dreidimensionale Struktur ergibt keine besonderen Auffälligkeiten, außer, dass im Falle der „aromatischen“ Hückelisomere das Molekül derart gefaltet ist, dass p-Orbitale durch den Raum in Kontakt geraten.

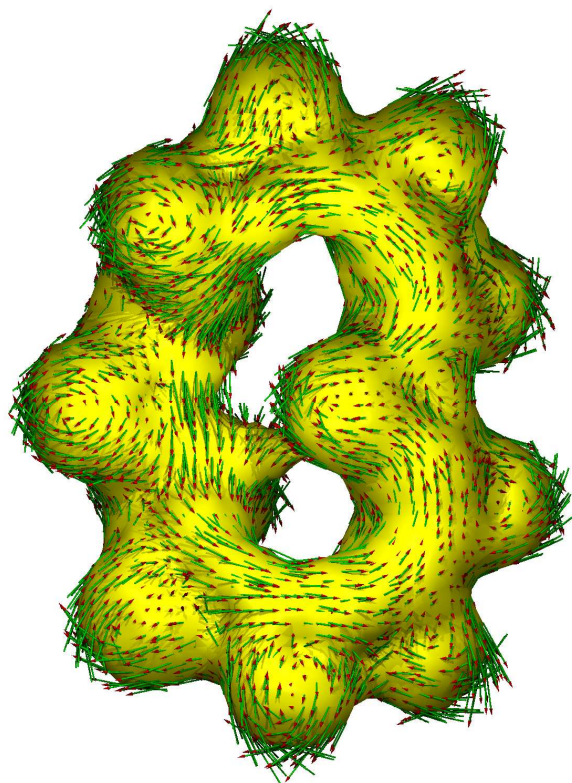
Die Graphen der Anisotropie der induzierten Stromdichte (AICD) zeigen bei einem *isosurface value* (IV-Wert) von ca. 0.1 .. 0.15 das etwaige Vorhandensein von *through-space* Konjugation an. In diesem Fall ist dann auch in Abbildung 189 der dazugehörige *critical isosurface value* (CIV-Wert) der transannularen Wechselwirkung angegeben.



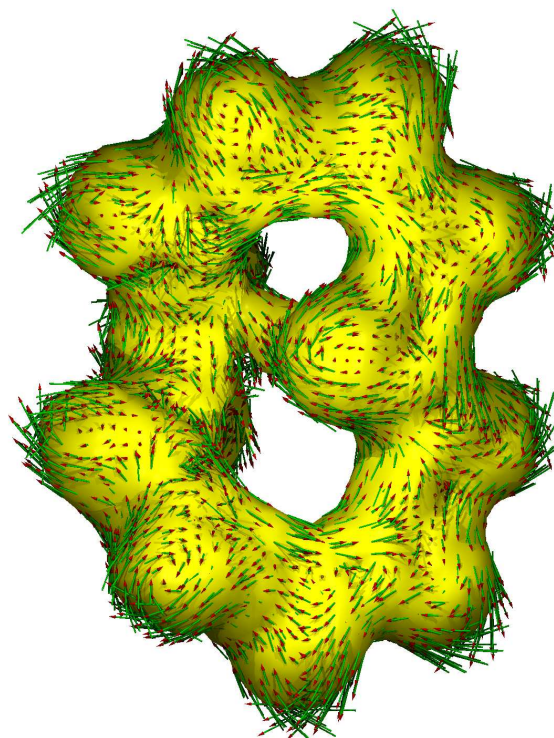
Hückel 1, CIV=0.012



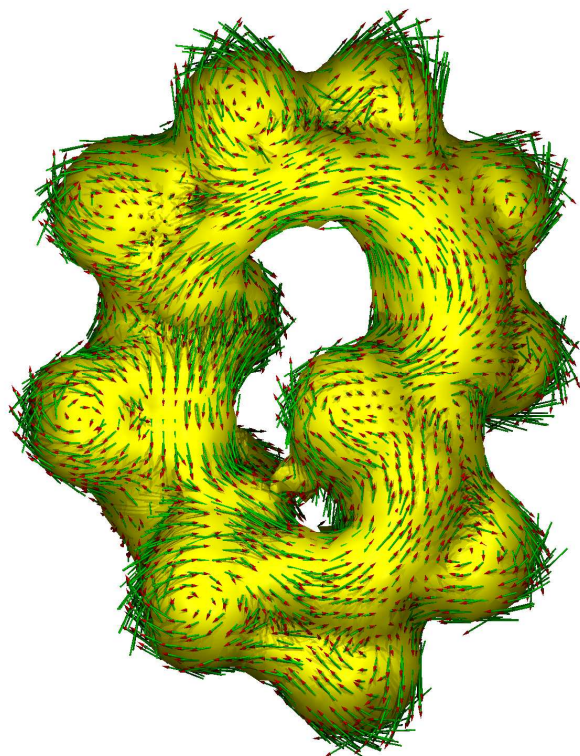
Hückel 61, CIV=0.013



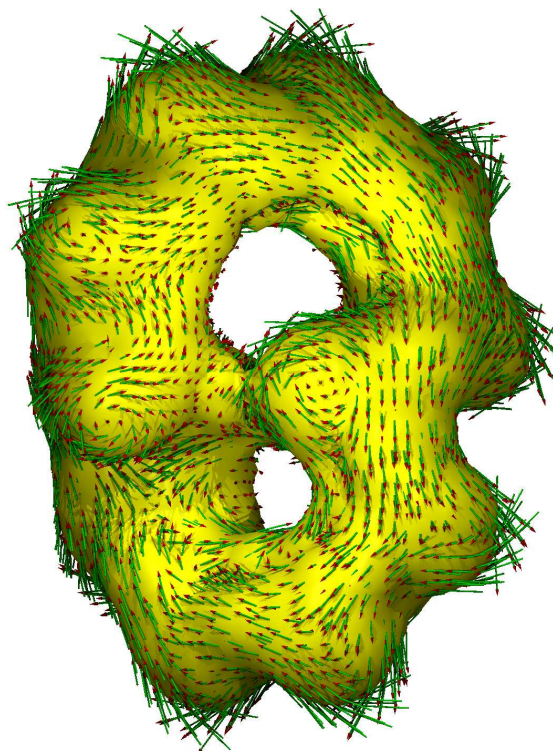
Hückel 121, CIV=0.016



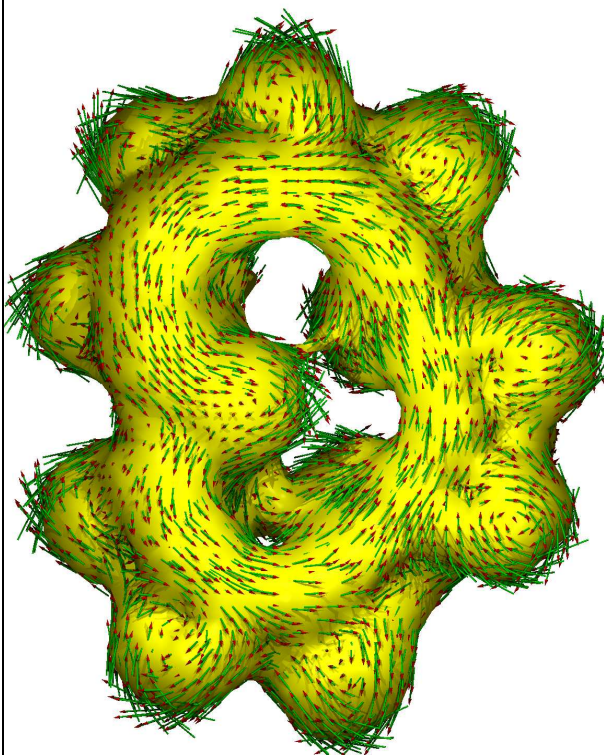
Hückel 146, CIV=0.021



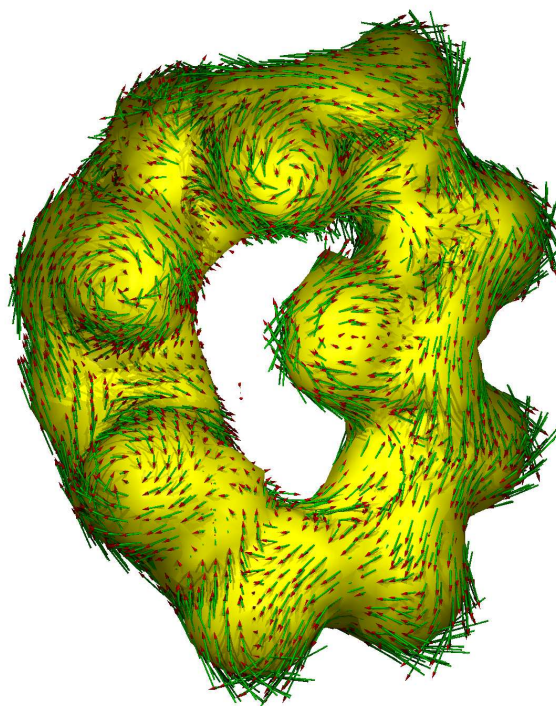
Hückel 218, CIV=0.016



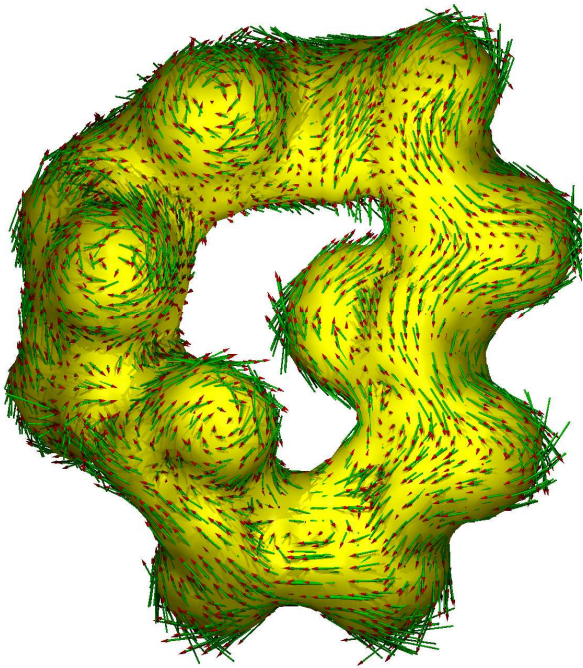
Möbius 421, CIV=0.024



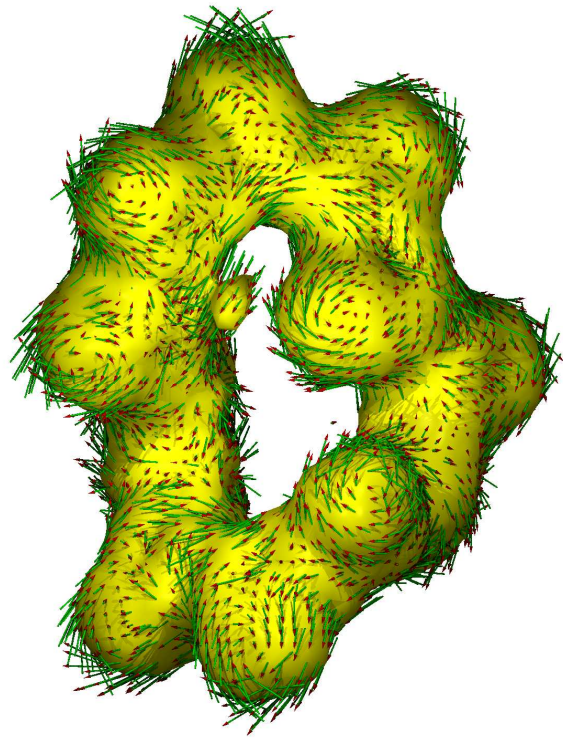
Hückel 1007, CIV=0.017



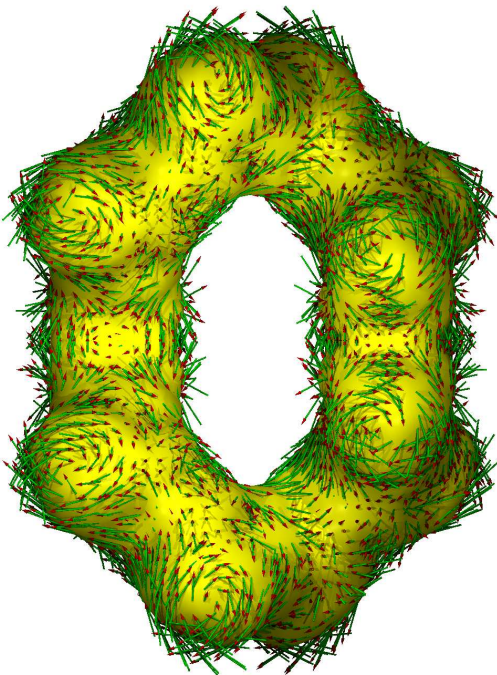
Möbius 1580, CIV=n/a



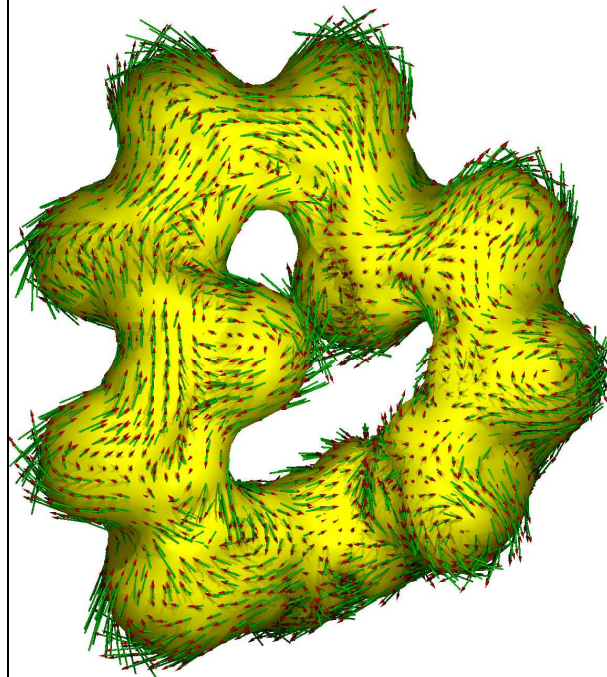
Möbius 2059, CIV=0.009



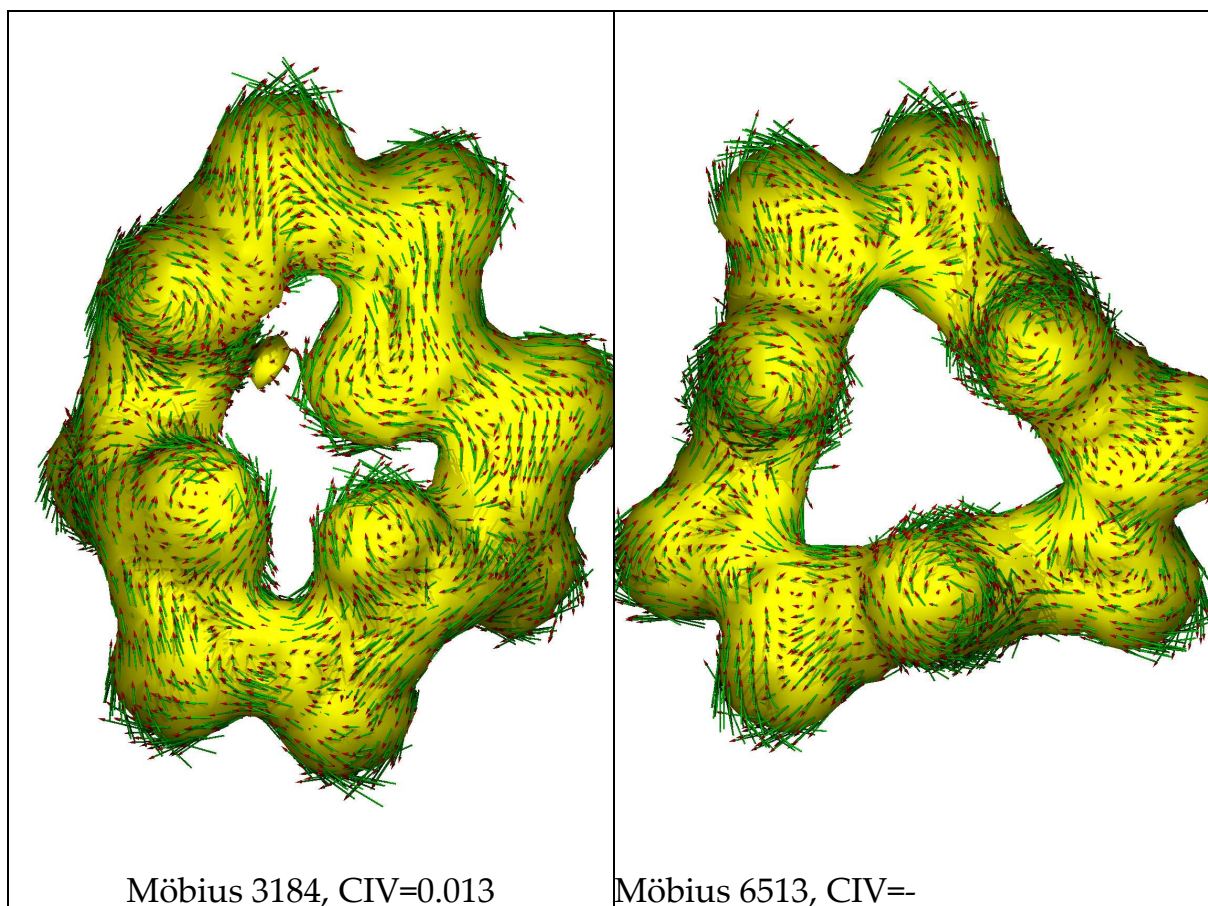
Hückel 2261, CIV=0.012



Hückel 2916, CIV=n/a



Möbius 2960, CIV=0.009



**Abbildung 189:** ACID-plots ausgesuchter B3LYP/6-31G\* optimierter  $[12]$ Annulen-Strukturen bei einem IV von 0.015 sowie CIV-Wert der *through-space* Konjugation

Aus den obigen Graphen lässt sich das unerwartet aromatische Verhalten der beiden erwähnten Isomere erklären. Beide besitzen anscheinend eine ausreichend starke stabilisierende *through-space* Konjugation um die antiaromatischen Effekte der Hückelkonjugation überzukompensieren. Im Falle der Möbiusstrukturen sind sämtliche elektronischen Effekte stabilisierend, so dass es zu keinen Ausreißern kommt.

### 3.6 $[16]$ Annulen

Der nächst höhere Vertreter in der Reihe  $[4n]$ Annulene ist das  $[16]$ Annulen. Ähnlich wie im  $[12]$ Annulen sind auch hier Möbius-aromatische Strukturen postuliert worden. Im folgenden Kapitel wird eine systematische Analyse der Aromatizität der stabilsten Isomere des  $[16]$ Annulen vorgestellt. Die Daten der kompletten KONAN-Analysen sind wiederum auf der DVD mit dem *supplementary material* zu finden.

### 3.6.1 MM+ Startstrukturen

Die KONAN-Analyse für das [16]Annulen verläuft gemäß Abbildung 172 anhand des Kurzschemas: Konformationssuche(MM+/HYPERCHEM) -> Semiempirische *single-points*(PM3) -> DFT(B3LYP/6-31G\*).

#### 3.6.1.1 Konformationssuche

Als Atomtypen werden C3 und H verwendet; der Ring wird mit alternierenden Doppelbindungen definiert. Die Konformationssuche analog Kapitel 3.5.1 ergibt nach 34331 untersuchten Strukturen eine Anzahl von 6946 Strukturen im Energiebereich bis 50 kcal/mol. Davon verbleiben nach CANE-Dublettenreduktion 377 Strukturen.

#### 3.6.1.2 PM3 und Optimierung (B3LYP/6-31G\*)

Diese werden wiederum nach PM3 *single-point*-Energie sortiert und die energetisch niedrigsten (bis zu einer relativen Energie von 30kcal/mol) werden einer Optimierung unterzogen. Es verbleiben 101 Strukturen in einem relativen Energiebereich von 20 kcal/mol. Interessant ist die Auftragung der Energien der B3LYP/6-31G\*-optimierten Strukturen gegenüber den zugehörigen Kraftfeld (MM+) bzw. semiempirischen Strukturen.

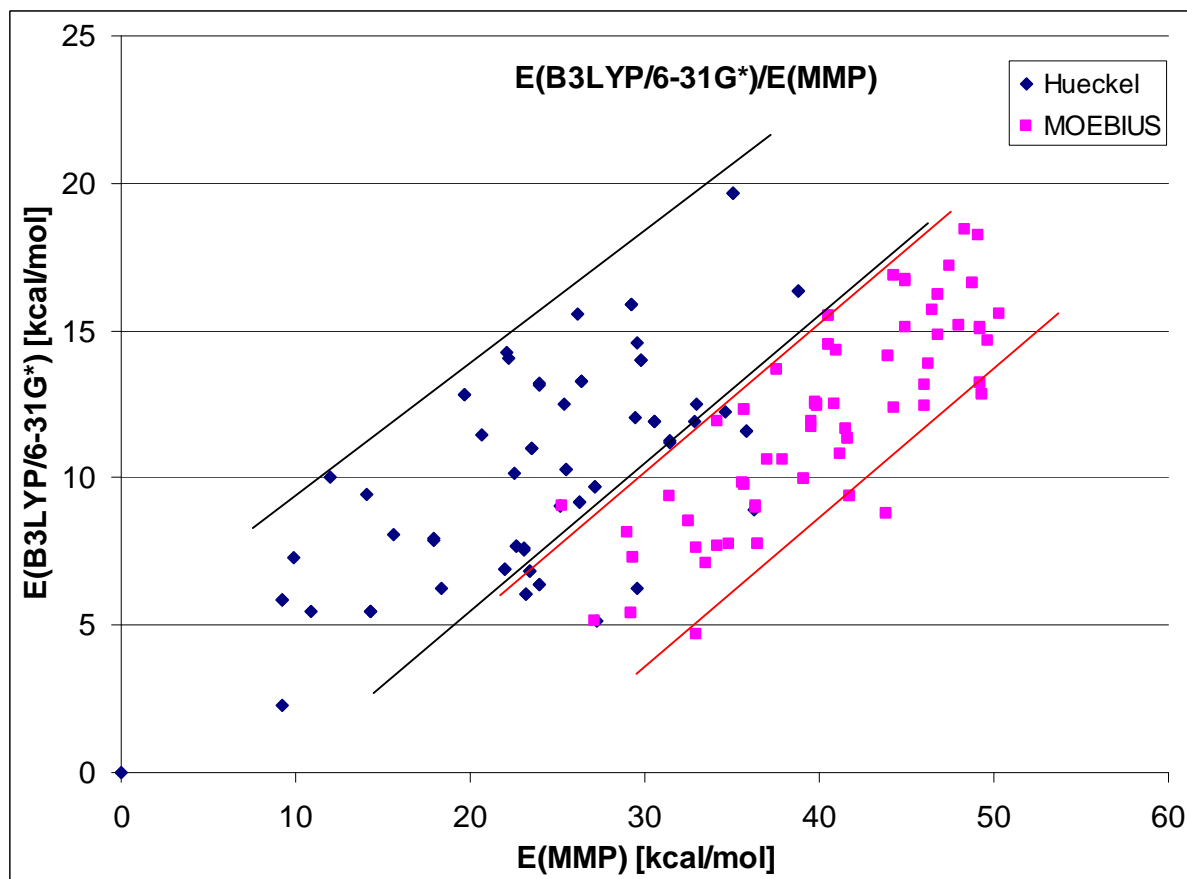


Abbildung 190: [16]Annulene, relative Energien der DFT-optimierten Strukturen über die Kraftfeldenergien der Strukturen aus der Konformationsuche

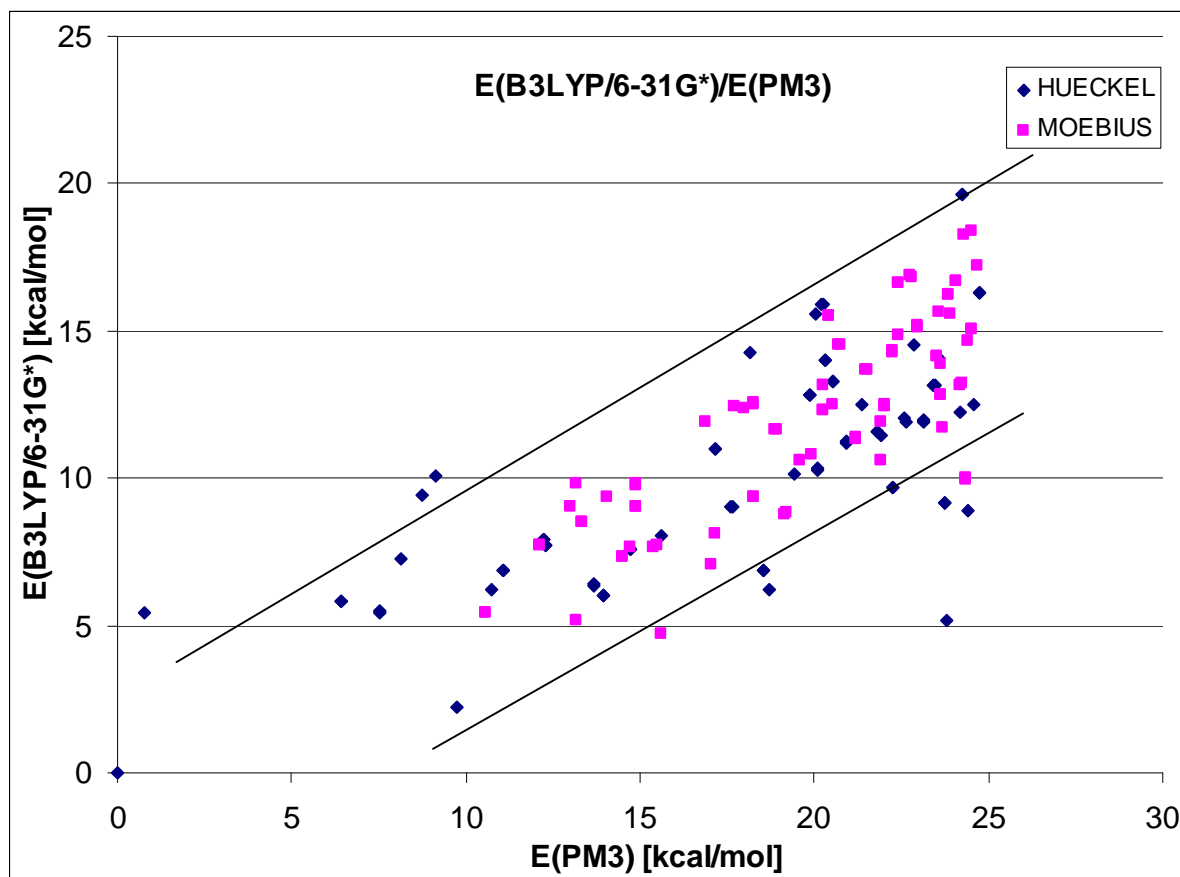


Abbildung 191: [16]Annulene, rel. Energien der DFT-optimierten Strukturen über der rel. PM3-Energien der Strukturen aus der Konformationsuche

Der elektronische Effekt der Möbiusaromatizität bzw. Hückelantiaromatizität wird durch Kraftfeldrechnungen nicht adäquat beschrieben. Dies zeigt sich in Abbildung 190 darin, dass die DFT-optimierten Möbius-Strukturen eine im Durchschnitt 8 kcal/mol niedrigere Energie aufweisen als die Hückelstrukturen bei einer Streuung von ca. 3-4 kcal/mol.

Bei Anlegen eines *energy-cutoff*, also eines Grenzwertes, oberhalb dessen Strukturen nicht mehr berücksichtigt werden, wären die Möbiusstrukturen damit systematisch benachteiligt. Dies würde bedeuten, dass wesentlich mehr Strukturen gerechnet werden müssten, um eine gleichwertige Abdeckung der Hyperfläche zu erzielen. Mit Hilfe des zwischengeschalteten PM3-Schrittes kann dies vermieden werden, da semiempirische Rechnungen sehr wohl elektronische Effekte berücksichtigen. Wie in Abbildung 191 zu erkennen ist, kann sowohl für Hückel- als auch für Möbiusisomere die zu erwartende DFT-Energie mit einer Streuung von ca.  $\pm 4$  kcal/mol abgeschätzt werden. Die Ausreißer in diesem Schema sind in der Regel durch Konfigurationsumwandlungen während der Optimierung zu erklären.

### 3.6.1.3 NICS/MAGS/FISE-Rechnungen

Ebenso wie in den Kapiteln 3.5.5 und 3.5.6 werden die NICS- und MAGS-Rechnungen für die DFT-optimierten Strukturen durchgeführt. Auf sämtliche Isomere mit geeigneten *cis*-Doppelbindungen wird die FISE-Methode angewandt. Die erhaltenen Daten sind in Tabelle 83 eingetragen.

**Tabelle 83:** [16]Annulen, die wichtigsten Aromatizitätskriterien aus der KONAN-Analyse nach Optimierung auf B3LYP/6-31G\* Niveau, Energien in kcal/mol

lognr	E <sub>rel</sub>	CANE string	T*	T <sub>max</sub> [°]	R <sub>av</sub> [Å]	A <sub>J</sub>	HOMA	NICS**	MAGS	FISE	FISE-NICS	FISE-MAGS
1	0.00	0011001100110011	H	32.6	1.4043	0.7570	0.3825	10.92	-59.67	3.85	9.97	26.55
61	2.25	0011011011001111	H	36.4	1.4060	0.7494	0.3489	10.43	-64.02	3.63	9.78	26.20
2223	4.68	0001100110011011	H	46.8	1.4049	0.7417	0.3423	6.39	-73.16	11.13	5.71	13.70
1255	5.14	0000110011000111	M	52.4	1.4038	0.8353	0.5642	-6.31	-103.66	14.63	-4.48	-14.27
1548	5.41	0001001100110011	M	51.3	1.4036	0.8367	0.5688	-6.18	-110.78	9.38	-6.18	-22.70
159	5.45	0011011011011011	H	31.2	1.4066	0.7495	0.3431	11.48	-61.10	11.51	10.10	27.83
260	5.45	0011011001101111	H	40.4	1.4060	0.7456	0.3406	8.93	-67.29	10.69	8.12	20.21
9	5.83	0001100110011011	H	43.9	1.4050	0.7464	0.3525	8.99	-66.73	5.73	8.34	23.28
736	6.05	0001010110011011	H	62.5	1.4063	0.7264	0.2942	1.40	-89.66	12.06	2.37	2.70
1641	6.22	0001001100110011	M	56.8	1.4039	0.8597	0.6184	-6.30	-112.72	13.26	-6.06	-23.41
407	6.24	0011001111001111	H	36.8	1.4062	0.7473	0.3422	9.42	-65.99	10.65	9.20	21.87
897	6.38	0001010011001111	H	45.9	1.4063	0.7498	0.3472	1.81	-86.17	11.25	2.92	3.77
791	6.84	0001011100011011	H	47.2	1.4059	0.7272	0.3000	0.86	-88.89	4.55	1.91	1.82
503	6.88	0010100110011011	H	42.9	1.4067	0.7369	0.3135	6.29	-76.74	11.62	5.41	10.08
2274	7.09	0001101100011111	M	45.2	1.4043	0.8085	0.4991	-5.22	-104.55	15.03	-5.80	-17.34
143	7.30	0001101100011011	H	15.9	1.4054	0.7651	0.3902	14.70	-46.33	12.48	13.64	35.83
1587	7.30	0001011000110011	M	47.5	1.4037	0.8356	0.5657	-6.19	-107.31	17.16	-6.44	-20.90
676	7.60	0011001101100111	M	51.6	1.4043	0.8678	0.6332	-7.46	-118.15	8.02	-7.45	-28.66



2162	7.62	0000101100110011	M	37.5	1.4036	0.8594	0.6200	-7.52	-113.37	9.84	-6.57	-23.45
2299	7.65	0011010110011011	M	39.4	1.4041	0.8751	0.6510	-8.31	-123.98	12.63	-7.92	-31.46
621	7.70	0001010011011011	H	46.7	1.4069	0.7415	0.3215	3.45	-82.46	3.98	4.21	8.47
2504	7.72	0000110110000111	M	47.6	1.4035	0.8855	0.6796	-10.26	-116.96	16.46	-8.34	-24.78
2841	7.73	0010110011011011	M	51.4	1.4051	0.8299	0.5398	-5.56	-113.61	15.10	-5.79	-21.02
353	7.90	0001111001101111	H	39.0	1.4063	0.7269	0.2951	7.42	-69.40	2.07	8.18	18.99
323	8.05	0001100110001111	H	48.0	1.4051	0.7377	0.3319	4.99	-77.65	14.48	5.99	13.17
1498	8.13	0001100011000111	M	47.3	1.4042	0.7991	0.4782	-4.00	-97.14	15.98	-4.63	-13.76
2059	8.50	0000101100011011	M	34.9	1.4034	0.8992	0.7116	-8.74	-120.49	12.63	-9.77	-36.73
4314	8.79	0010110011110011	M	58.3	1.4054	0.7927	0.4531	-4.63	-106.57	11.85	-4.29	-16.06
2758	8.89	0001111100011111	H	46.3	1.4060	0.7127	0.2662	-0.73	-92.41	10.20	2.13	-0.39
1007	9.02	0001010110001111	H	47.5	1.4063	0.7271	0.2954	0.24	-86.52	8.53	1.03	1.59
2764	9.04	0010110110011011	M	55.9	1.4052	0.8015	0.4754	-4.10	-106.87	8.70	-4.70	-18.01
1056	9.05	0000100001100011	M	45.8	1.4042	0.8163	0.5176	-4.05	-93.16	10.46	-2.99	-8.92
1156	9.19	0001110011001111	M	46.1	1.4047	0.8133	0.5057	-5.41	-104.23	6.41	-7.88	-22.09
3908	9.37	0001100100110011	M	55.4	1.4044	0.8023	0.4840	-4.30	-106.17	16.27	-4.22	-15.32
1875	9.39	0000010000110011	M	35.6	1.4028	0.9444	0.8186	-10.79	-122.60	17.41	-6.44	-27.34
254	9.45	0001101100011011	H	10.1	1.4060	0.7715	0.3993	19.01	-39.71	7.54	16.40	40.88
1289	9.71	0000101110001111	H	44.4	1.4059	0.7346	0.3169	0.39	-88.38	10.54	3.26	3.51
2598	9.77	0001001100011011	M	54.8	1.4043	0.8170	0.5185	-5.39	-109.32	10.22	-5.39	-18.42
2568	9.83	0000100110011011	M	38.6	1.4039	0.8703	0.6418	-6.96	-114.34	14.87	-6.94	-25.26
3240	9.96	0001100110011111	M	52.1	1.4045	0.7825	0.4382	-4.93	-103.72	7.35	-4.20	-14.22
222	10.05	0001100110110011	H	18.2	1.4056	0.7625	0.3824	15.51	-46.27	10.45	14.38	37.77
595	10.15	0001101101100111	M	50.5	1.4052	0.8390	0.5598			9.77	-6.08	-20.21
1085	10.31	0000101000110011	H	64.3	1.4058	0.7259	0.2981	0.55	-87.44	13.86	1.90	2.34
2991	10.60	0001010100011011	M	47.6	1.4059	0.7352	0.3183	-2.49	-96.81	14.35	-2.53	-8.16
3170	10.61	0001011001100011	M	50.8	1.4044	0.7920	0.4609	-4.78	-103.23	8.94	-4.23	-13.72
3737	10.78	0010011011001111	M	69.0	1.4058	0.7452	0.3413	-2.03	-98.74	13.14	-2.21	-9.40
835	11.00	0001101111001111	H	38.3	1.4066	0.7406	0.3232	8.48	-65.79	10.25	8.93	20.89
1827	11.24	0001101010011011	H	57.6	1.4067	0.7083	0.2487	2.59	-83.70	6.92	2.54	4.56
3881	11.34	0001001101100011	M	64.2	1.4051	0.7644	0.3922	-3.65	-101.41	8.69	-3.99	-13.29
480	11.47	0001111000111111	H	41.7	1.4058	0.7369	0.3234	4.56	-77.65	11.35	7.43	12.75
2625	11.59	0001010001010011	H	48.7	1.4068	0.7376	0.3139	-1.33	-92.04	13.22	0.90	-1.07
3858	11.64	0000100110110011	M	57.3	1.4051	0.7952	0.4617	-4.36	-104.88	13.40	-3.85	-13.65
3303	11.69	0011011001111111	M	57.9	1.4046	0.8100	0.4998	-6.54	-112.36	19.58	-5.84	-20.32
2321	11.89	0000010001100011	M	40.8	1.4041	0.8639	0.6262	-5.83	-102.13	15.75	-3.16	-11.50
1749	11.92	0000010011000111	H	42.6	1.4061	0.7383	0.3231	-1.67	-91.83	11.97	0.75	-2.16
2116	11.93	0000101011001111	H	68.8	1.4069	0.7035	0.2352	0.50	-88.68	10.41	1.17	-0.58
3281	11.94	0000100011001111	M	81.3	1.4064	0.7193	0.2765	-1.97	-93.87	3.18	-1.78	-6.07
1618	12.02	0001010001100011	H	44.5	1.4055	0.7322	0.3147	0.92	-87.52	6.04	2.51	4.20
2440	12.24	0000111001100111	H	50.6	1.4066	0.7129	0.2606	0.05	-89.02	4.63	1.49	0.64
2586	12.33	0000100011000011	M	61.8	1.4053	0.7487	0.3541	-2.55	-91.63	14.29	-1.98	-6.37
4610	12.34	0001110011001111	M	65.8	1.4054	0.7785	0.4205	-3.64	-101.54	12.19	-3.33	-12.17
5089	12.44	0010011011011011	M	67.2	1.4064	0.7555	0.3584	-2.09	-102.41	10.90	-2.78	-10.69
3481	12.47	0000110001100111	M	47.0	1.4040	0.8603	0.6189	-8.26	-110.02	15.46	-6.22	-19.68
2191	12.48	0000101100011111	H	44.9	1.4062	0.7053	0.2471	0.72	-88.67	11.75	2.75	1.96
3666	12.49	0000100110001111	M	45.7	1.4044	0.8171	0.5171	-5.47	-105.45	14.69	-6.58	-19.45
1067	12.49	0000101000011011	H	54.4	1.4063	0.7377	0.3195	2.51	-81.21	12.78	5.08	8.87
3395	12.55	0010110110110011	M	60.4	1.4058	0.7972	0.4591	-4.40	-107.53	14.82	-4.79	-17.66
439	12.82	0000110110001111	H	34.2	1.4055	0.7618	0.3817	8.95	-65.22	15.63	8.51	17.25
6467	12.83	0010110011001111	M	67.2	1.4061	0.7595	0.3714	-3.26	-101.60			
5061	13.13	0001010011001111	H	63.4	1.4071	0.7050	0.2373	3.21	-81.86	12.28	3.35	6.54
951	13.18	0001101100111111	H	42.0	1.4059	0.7531	0.3581	7.36	-71.84	4.00	8.50	20.06
6411	13.19	0001010011001011	M	60.6	1.4066	0.7285	0.2951	-2.52	-96.18			
1206	13.27	0001010001101111	H	45.4	1.4065	0.7554	0.3575	1.37	-83.08	11.78	4.73	7.41
3025	13.65	0001011110011011	M	49.6	1.4044	0.8524	0.5973	-7.63	-113.86	13.63	-7.05	-24.66
5146	13.86	0001010010011011	M	77.0	1.4071	0.6893	0.2011	-0.63	-92.11	5.64	-0.56	-2.79

1720	13.98	0000100100011011	H	45.4	1.4066	0.7307	0.3001	-0.56	-91.26	4.73	0.50	-1.14
586	14.04	0000111100001111	H	28.8	1.4054	0.7655	0.3912	10.40	-68.01	9.56	11.14	17.99
4414	14.13	0010110011001111	M	58.6	1.4056	0.7957	0.4578	-4.01	-103.56	15.82	-4.00	-14.35
563	14.26	0000110110011011	H	76.5	1.4060	0.7242	0.2925	3.62	-81.55	3.62	3.76	7.22
3685	14.30	0011001101011011	M	43.8	1.4046	0.8624	0.6182	-7.49	-120.80	16.99	-7.19	-29.16
3524	14.53	0000110100011011	M	50.5	1.4045	0.8219	0.5275	-6.79	-106.63	15.47	-6.13	-19.35
1692	14.56	0000010110001101	H	44.6	1.4062	0.7303	0.3039	-3.63	-94.29	13.13	-0.92	-6.39
6602	14.63	0011011110011111	M	53.1	1.4065	0.7810	0.4153	-5.21	-106.40			
5486	14.83	0001011001101111	M	59.0	1.4063	0.7640	0.3795	-4.51	-103.76			
6433	15.07	0010010100110011	M	61.2	1.4067	0.7226	0.2808	-1.47	-93.49			
4794	15.10	0000101001100111	M	50.3	1.4060	0.7587	0.3698	-4.50	-97.23	12.34	-3.75	-7.86
5924	15.16	0000010110011011	M	58.4	1.4051	0.8195	0.5165	-6.88	-109.18			
3558	15.50	0001011011011011	M	55.2	1.4061	0.7892	0.4384	-4.17	-105.52	8.50	-4.54	-16.71
1127	15.55	0011001101111011	H	54.5	1.4064	0.7475	0.3407	9.39	-64.47	2.83	9.02	24.46
6847	15.55	0001011001111011	M	72.2	1.4066	0.7559	0.3581	-3.32	-100.77			
5295	15.66	0000010000011011	M	29.3	1.4028	0.9964	0.9352	-14.76	-169.60			
1513	15.89	0001101101101111	H	38.8	1.4073	0.7530	0.3443	8.32	-69.78	13.52	7.97	19.83
5459	16.21	0001001010011011	M	62.5	1.4069	0.7264	0.2871	-2.10	-93.43			
3220	16.32	0001100101001111	H	49.7	1.4070	0.7263	0.2862	3.41	-82.38	14.88	3.82	5.32
6190	16.61	0000100110001111	M	66.9	1.4050	0.8256	0.5315	-5.50	-106.78			
4754	16.69	0001101011001111	M	41.2	1.4046	0.8942	0.6896	-10.10	-128.40	8.38	-8.14	-33.08
4638	16.86	0001101111000111	M	49.7	1.4060	0.8019	0.4681	-4.96	-101.13	12.64	-5.04	-14.07
5702	17.21	0000010011001111	M	52.7	1.4050	0.8320	0.5455	-7.16	-107.32			
6307	18.24	0000010011000101	M	50.8	1.4067	0.7345	0.3085	-3.12	-93.98			
5998	18.41	0000101100001111	M	62.0	1.4045	0.8536	0.5989	-7.20	-110.97			
2512	19.66	0000101000110101	H	60.2	1.4076	0.6636	0.1374	-1.07	-87.81	9.92	1.87	2.51

\*Topology: H=Hückel, M=Möbius

\*\*siehe Anmerkung zu Tabelle 81

### 3.6.1.4 Strukturelle Kriterien

Auch für das [16]Annulen zeigen HOMA und der Julg-Index  $A_j$  eine lineare Abhängigkeit, siehe Abbildung 192.

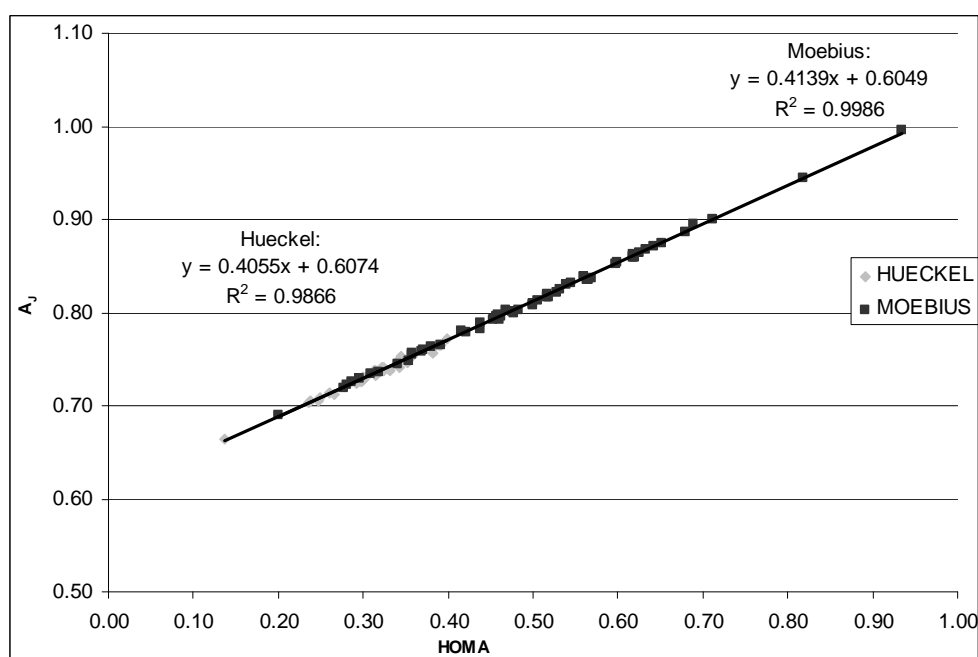


Abbildung 192: [16]Annulen, Julg-Index über HOMA

Im Vergleich zum [12]Annulen (siehe Abbildung 178) liegen die Regressionskoeffizienten noch deutlich näher bei eins, was wahrscheinlich auf die geringere Spannung durch den größeren Ring zurückzuführen ist. Die Verteilung der Isomere über die strukturellen Kriterien (hier exemplarisch für HOMA) ist in Abbildung 193 wiedergegeben.

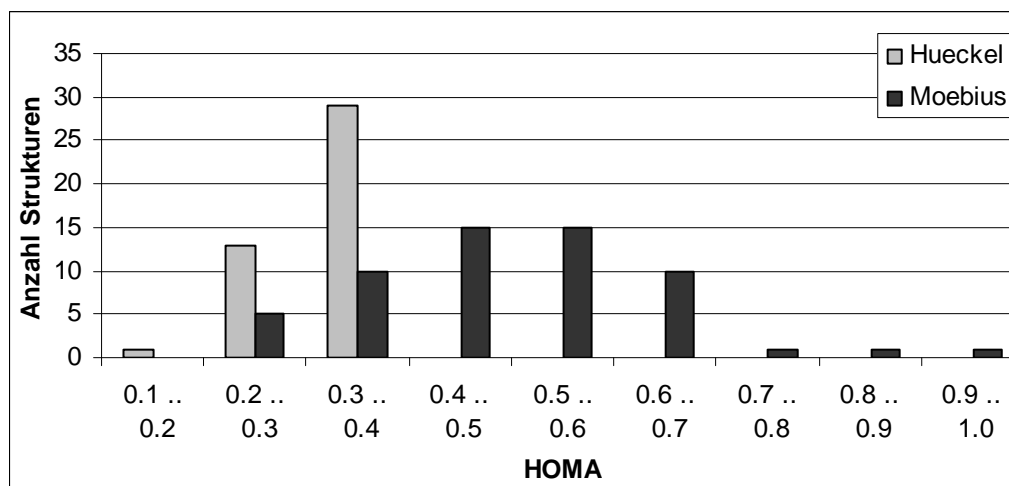


Abbildung 193: [16]Annulen, Histogramm der Verteilung der HOMA-Werte

Im Gegensatz zum [12]Annulen (Abbildung 179) zeigen die [16]Annulen-Isomere strukturell deutlich aromatischeres Verhalten.

### 3.6.1.5 Energetische Kriterien

Die Verteilung der relativen B3LYP/6-31G\*-Energie ist als Histogramm in Abbildung 194 aufgetragen.

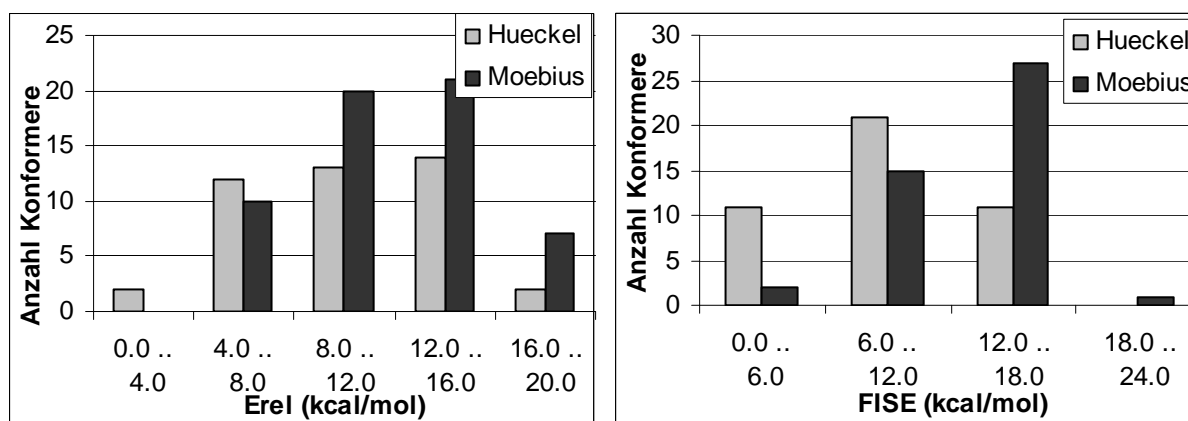


Abbildung 194: [16]Annulen, Histogramme  $E_{rel}$  links und FISE rechts

In der Gesamtenergie scheinen Möbiusisomere nicht besonders bevorzugt. Bei der FISE-Energie liegt das Maximum der Verteilung für Hückelisomere bei ca. 12 kcal/mol ( $\bar{\varnothing}$ =9.4 kcal/mol). Für Möbiusisomere liegt das Verteilungsmaximum bei 15 kcal/mol ( $\bar{\varnothing}$ =12.5 kcal/mol). Die FISE-Energie als Konjugationsstabilisierungsenergie (CSE) ist für beide Gruppen, Hückel- und Möbiusisomere

positiv, wird also augenscheinlich durch lineare Konjugation und *through-space*-Konjugation dominiert. Die Möbiusisomere sind im Schnitt ca. 3 kcal/mol stabiler, was dem Einfluß der aromatischen Stabilisierung entspricht. Die aromatische Stabilisierung liegt damit noch innerhalb der Streuung der FISE-Energie. Bei den  $[4n]$ Annulenen ist die Aromatizität also ein deutlich schwächerer Einflussfaktor als z.B. bei den planaren Aromaten, wo die Aromatizität die Struktur der Moleküle vollständig bestimmt, siehe Kapitel 1.4.3.

Leider lassen sich die absoluten Zahlen für die FISE-Energien bei den  $[4n]$ Annulenen aufgrund des Einflusses von linearer und *through-space* Konjugation schlecht interpretieren. Es müssen immer relative Werte im Rahmen eines Isomerenensembles betrachtet werden, um zu einer Aussage zu kommen. Die Auftragung der FISE-Werte über die HOMA-Werte in Abbildung 195 ergibt qualitativ das gleiche Bild wie schon beim  $[12]$ Annulen.

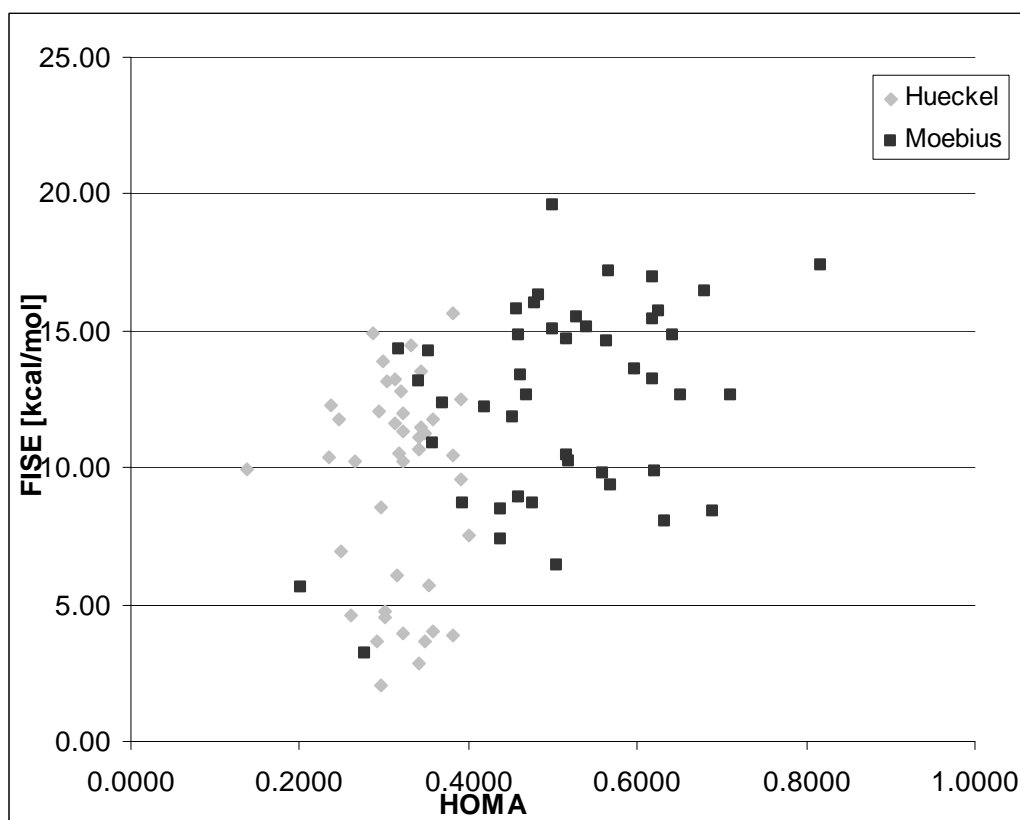
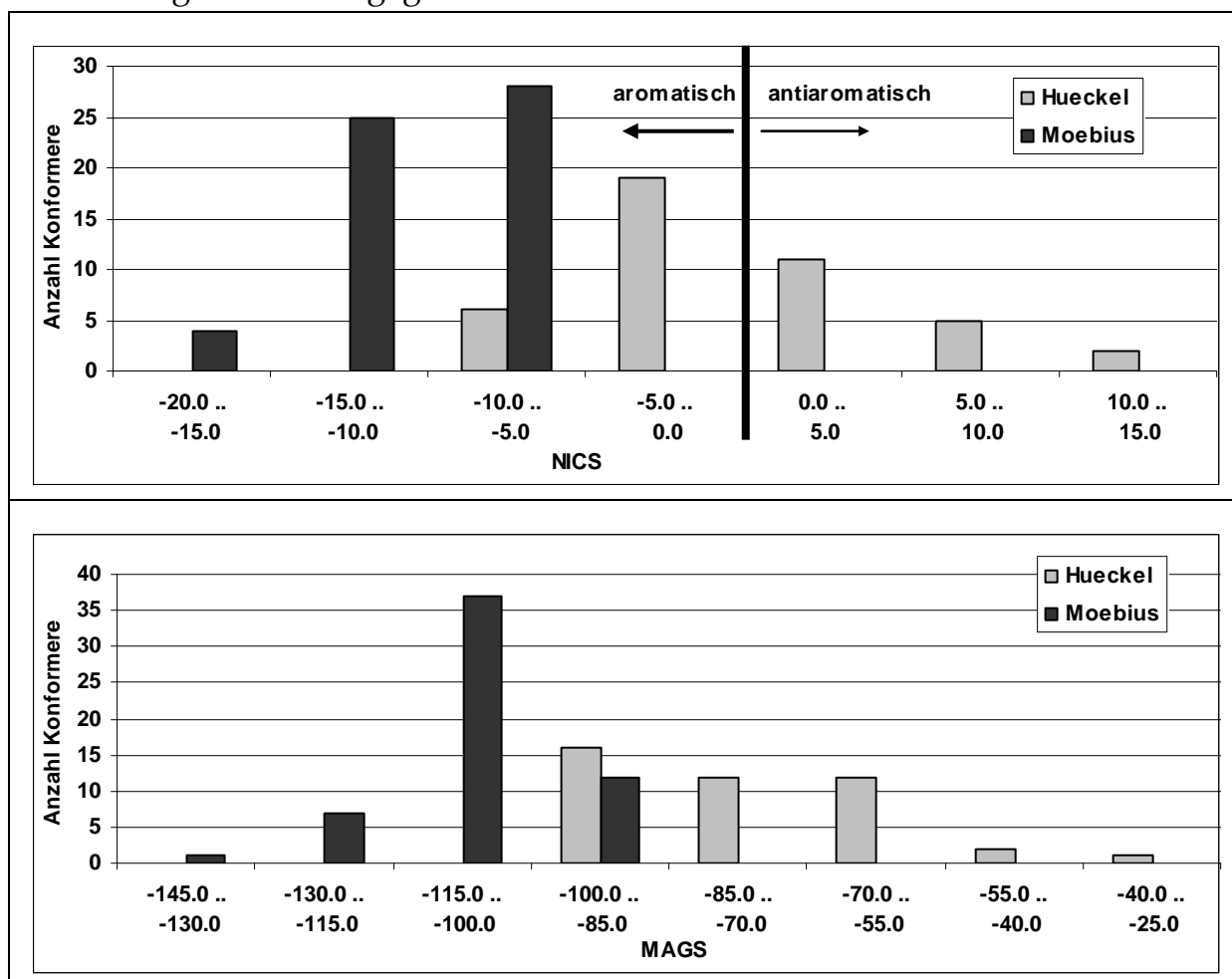


Abbildung 195:  $[16]$ Annulene, FISE über HOMA

Mit großer Streuung kann ansatzweise von einem linearen Zusammenhang des energetischen Kriteriums mit dem strukturellen Kriterium gesprochen werden. Interessant ist dies insofern, als dass für die antiaromatischen Hückelisomere eigentlich von einem reziproken Zusammenhang zwischen Delokalisierung und Energie ausgegangen werden sollte. Offenbar überkompensiert der Energiegewinn durch einfache Delokalisierung den aromatischen Term im Falle der Hückelisomere.

### 3.6.1.6 Magnetische Kriterien

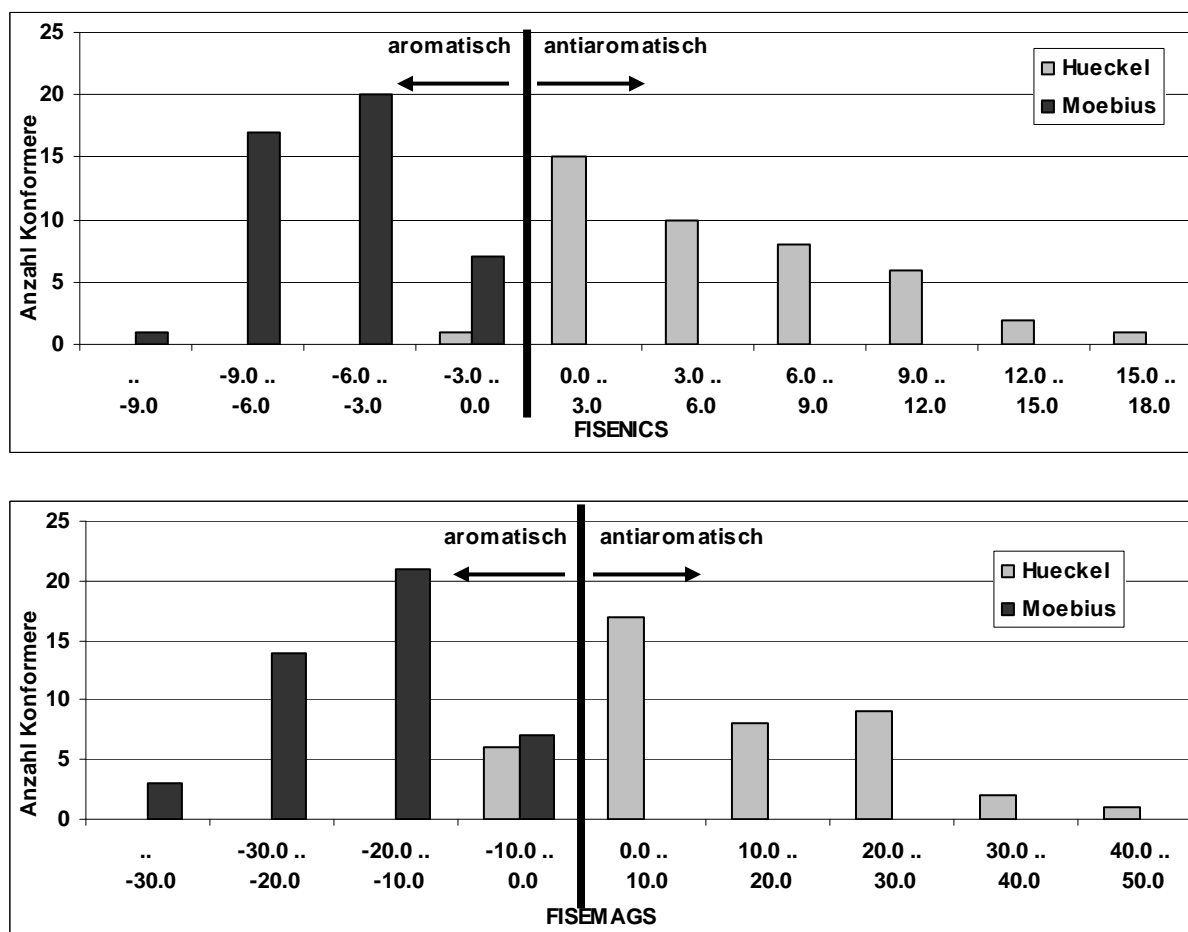
Die Verteilung der Werte für die traditionellen magnetischen Kriterien ist in Abbildung 196 wiedergegeben.



**Abbildung 196: [16]Annulen, Histogramme NICS (links) und MAGS (rechts)**

Im Gegensatz zu den [12]Annulenen lassen sich die [16]Annulen-Isomere bereits anhand der klassischen magnetischen Kriterien recht gut in aromatisch und nichtaromatisch klassifizieren. Der Vollständigkeit halber sind auch die Verteilung der FISENICS und FISEMAGS-Werte als Histogramm ausgewertet und in Abbildung 197 dargestellt.

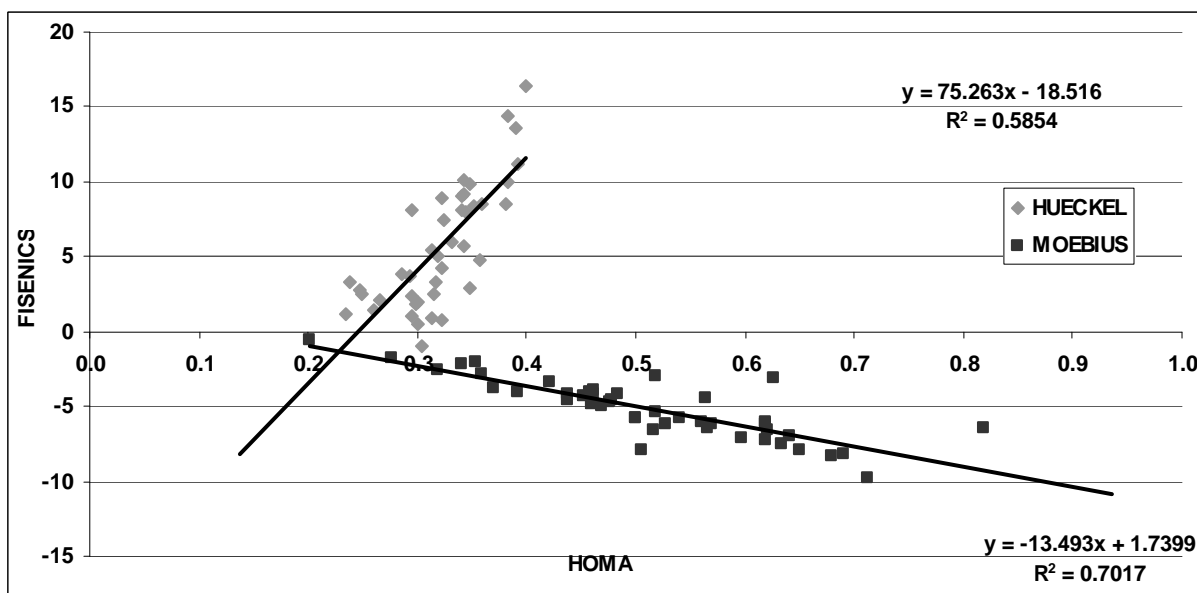
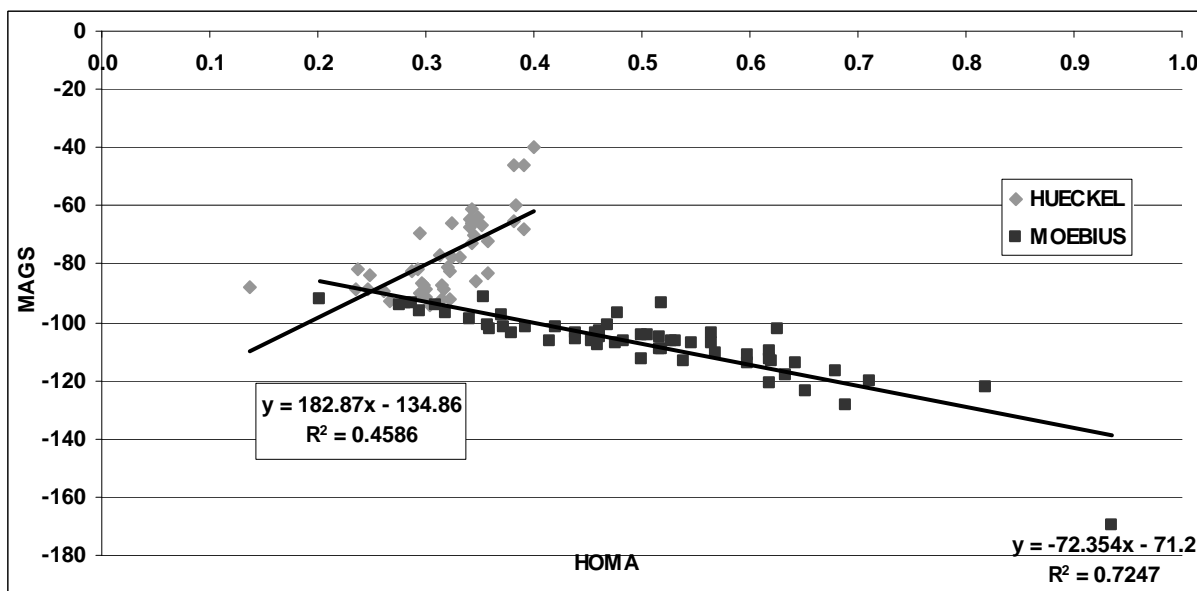
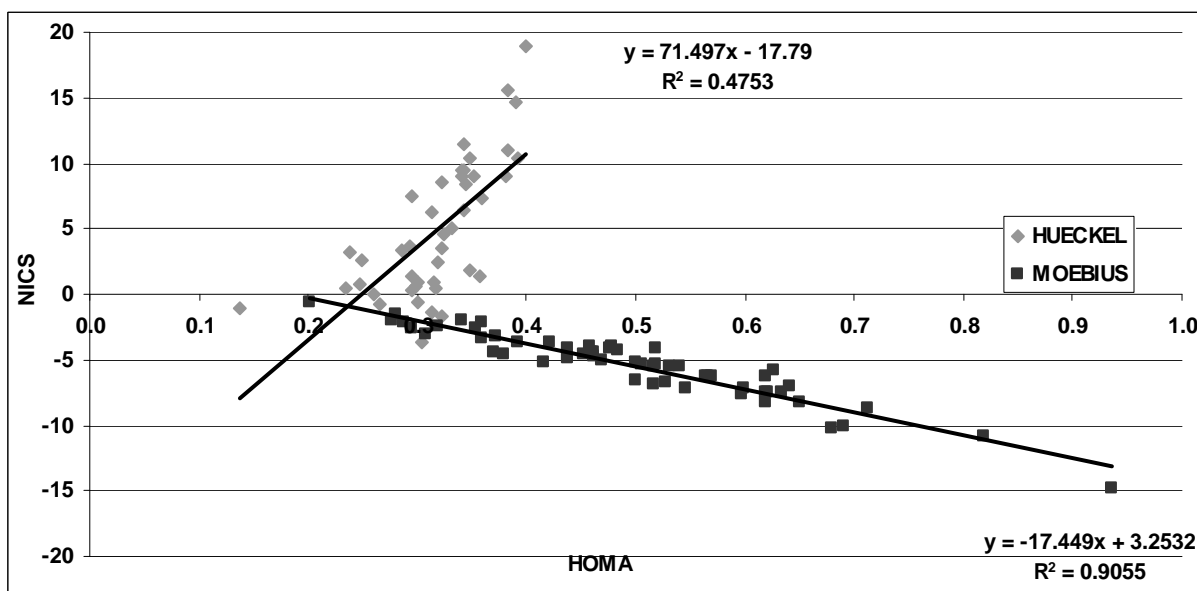
Wiederum zeigt sich die klarere Unterscheidung im Vergleich zu den NICS- und MAGS-Werten. Im Falle der FISEMAGS-Werte kommt als weiterer Vorteil im Vergleich zu den MAGS-Werten die mögliche absolute Verwendung einzelner Werte ohne relativen Bezug zu vergleichbaren Strukturen.



**Abbildung 197: [16]Annulen, Histogramme FISENICS (links) und FISEMAGS (rechts)**

Um die Korrelation der magnetischen Aromatizitätskriterien mit den strukturellen zu untersuchen, werden diese über die HOMA-Werte aufgetragen in Abbildung 198. Wieder zeigt sich eine klare Trennung zwischen den Hückel- und Möbiusisomeren. Bei Letzteren ist die „magnetische Aromatizität“ mit der „strukturellen Aromatizität“ positiv korreliert. Dies ist typisch für klassische Aromaten wie Benzol. Bei den Hückelisomeren gibt es eine negative Korrelation von „magnetischer Aromatizität“ und „struktureller Aromatizität“, wie es typisch ist für klassische Antiaromaten. Der Grund dafür liegt in der Überlappung des  $\pi$ -Systems. Diese ist sowohl für Aromaten als auch Antiaromaten direkt mit den strukturellen Aromatizitätskriterien korreliert. Aromaten mit großer Überlappung des  $\pi$ -Systems sind in ihren aromatischen Eigenschaften besonders ausgeprägt, Antiaromaten entsprechend besonders antiaromatisch.

Auffällig ist des Weiteren der größere FISENICS- und FISEMAGS-Wertebereich, der von den Hückelisomeren populierte wird. Dieses ist wahrscheinlich auf die konkurrierenden Effekte der „antiaromatischen“ und linearen sowie *through-space* elektronischen Konjugation zurückzuführen.



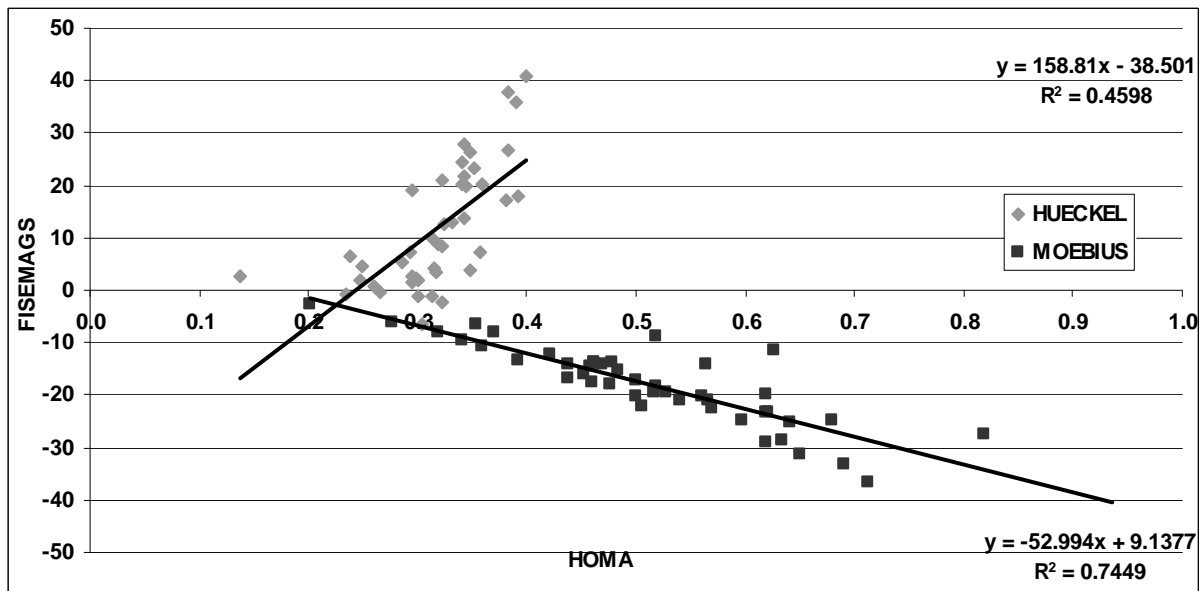


Abbildung 198:  $[16]$ Annulen – Korrelation der magnetischen Aromatizitätskriterien mit den HOMA-Werte

Die Korrelation der magnetischen Aromatizitätskriterien mit den energetischen wird exemplarisch in Abbildung 199 untersucht.



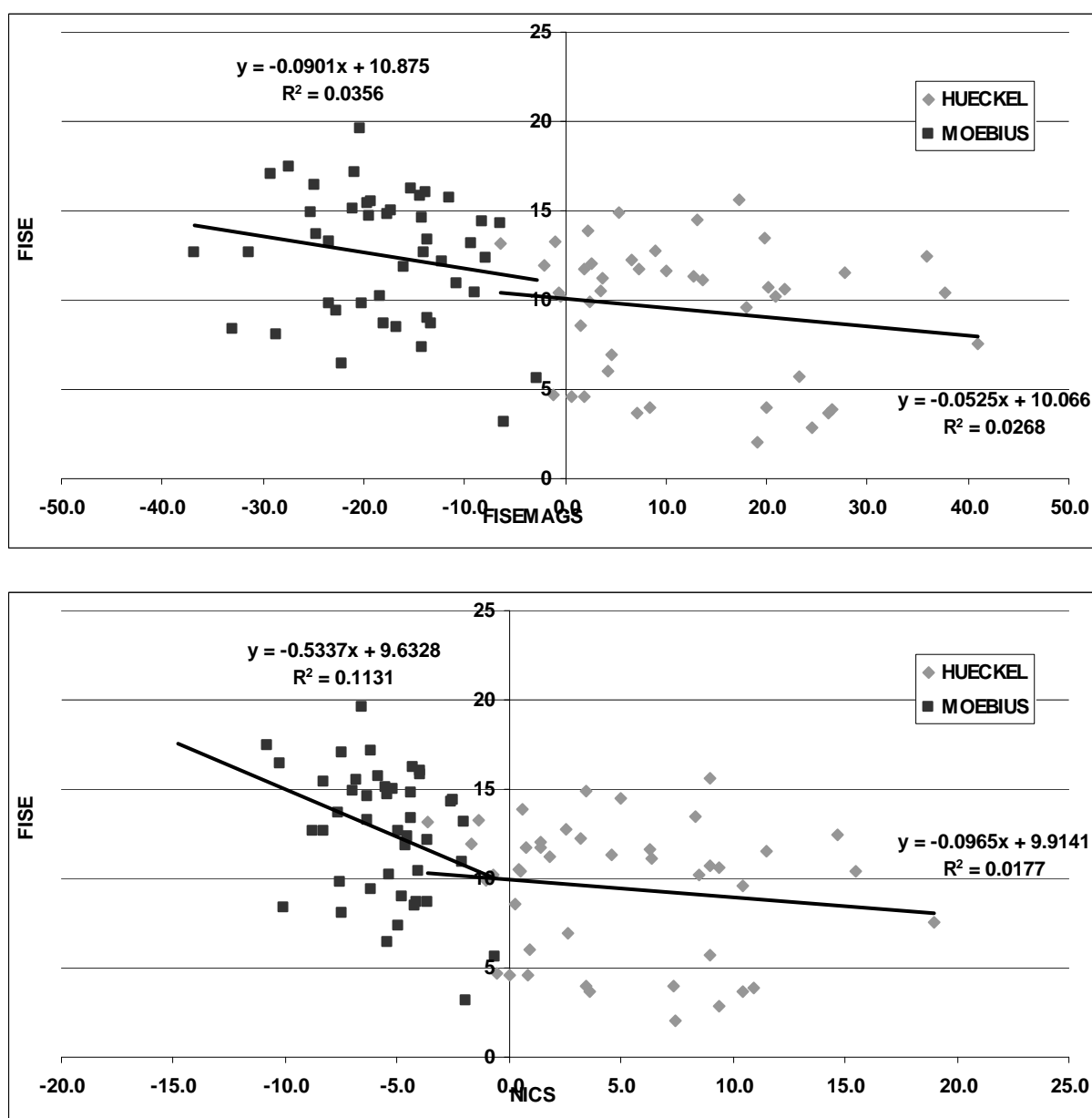


Abbildung 199: [16]Annulen, Korrelation der FISE-Energie mit einem der magnetischen Kriterien

Obwohl unbestreitbar ein wichtiges Kriterium, zeigt die FISE-Energie nur eine geringe Korrelation zu den magnetischen Kriterien, wobei sich wieder die Hückel- und Möbiusisomere unterscheiden.

### 3.6.1.7 Principal Component Analysis (PCA)

Im folgenden Abschnitt wird eine Hauptkomponentenanalyse unter Verwendung der Korrelationsmatrix mit den verschiedenen Aromatizitätskriterien als Eingabedimensionen durchgeführt. Verwendet wird das Plugin OOSTAT Version 0.4<sup>[131]</sup> für OPENOFFICE 2.3.1<sup>[132]</sup> Die Korrelationsmatrix für alle Isomere, für die ein vollständiger Satz an Daten erhältlich ist, wird in Tabelle 84 angegeben.

**Tabelle 84: [16]Annulen – Korrelationsmatrix Aromatizitätskriterien**

Correlation	$T_{\max}$	HOMA	$A_J$	FISE	NICS	MAGS	FISE-NICS	FISE-MAGS
$T_{\max}$	1	-0.22	-0.22	-0.03	-0.43	-0.41	-0.43	-0.39
HOMA	-0.22	1	1	0.44	-0.62	-0.66	-0.65	-0.69
$A_J$	-0.22	1	1	0.44	-0.62	-0.67	-0.65	-0.69
FISE	-0.03	0.44	0.44	1	-0.41	-0.4	-0.4	-0.42
NICS	-0.43	-0.62	-0.62	-0.41	1	0.98	0.98	0.98
MAGS	-0.41	-0.66	-0.67	-0.4	0.98	1	0.98	0.99
FISENICS	-0.43	-0.65	-0.65	-0.4	0.98	0.98	1	0.99
FISEMAGS	-0.39	-0.69	-0.69	-0.42	0.98	0.99	0.99	1

Aus der Korrelationsmatrix ist zu erkennen, dass die magnetischen Kriterien und die strukturellen Kriterien stark korreliert sind. Die maximale Abweichung des Diederwinkels aus der Planarität ( $T_{\max}$ ) ist eher mit den magnetischen Kriterien korreliert als mit den strukturellen. Die FISE-Energie als energetisches Kriterium ist etwa gleich stark mit den magnetischen und den strukturellen Kriterien korreliert. Die Reduktion der Aromatizitätskriterien auf möglichst wenige signifikante Parameter ist Thema des folgenden Kapitels. In obiger Tabelle sind die praktisch perfekt miteinander korrelierten Parameter als graue Blöcke gekennzeichnet. Prinzipiell kann der Informationsgehalt der Datensätze durch nur vier Einflussgrößen beschrieben werden, wobei je eine Einflussgröße aus jedem der vier Blöcke gewählt wird.

Eine PCA-Analyse mit Hilfe der Korrelationskoeffizientenmatrix  $\Sigma'$  (Gleichung (1.25)) ergibt die Hauptkomponenten in Tabelle 85. Bereits mit zwei Hauptkomponenten werden 87% der Gesamtvarianz erklärt. Mehr als 95% der Gesamtvarianz werden durch drei Hauptkomponenten repräsentiert. Im Folgenden werden diese drei Hauptkomponenten genauer untersucht.

**Tabelle 85: [16]Annulen, Hauptkomponenten, ihre Eigenwerte und Anteile an der Gesamtvarianz**

	PC_1	PC_2	PC_3	PC_4	PC_5	PC_6	PC_7	PC_8
Eigenwert	5.3766	1.6438	0.7241	0.2097	0.0232	0.0187	0.0032	0.0007
%Varianz	67.2071	20.5479	9.0511	2.6210	0.2901	0.2340	0.0399	0.0088
%Varianz, kumulativ	67.21	87.76	96.81	99.43	99.72	99.95	99.99	100.00

$\Gamma$ , die Matrix der Eigenvektoren  $\gamma_i$  (1.23), ist in Tabelle 86 wiedergegeben.

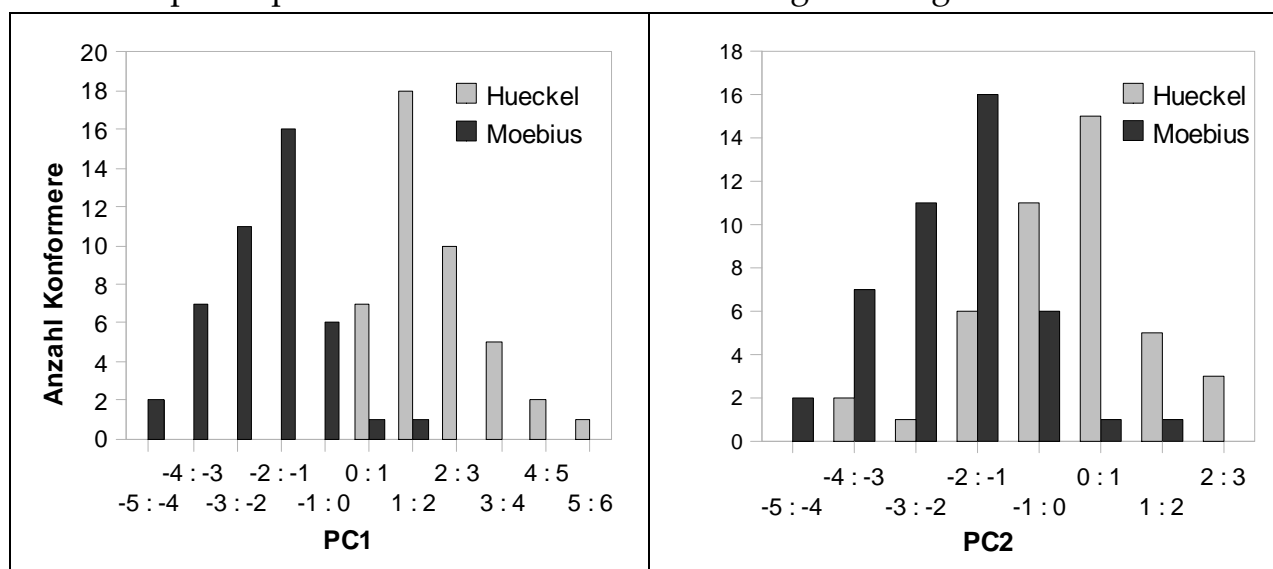
**Tabelle 86:** [16]Annulen, Eigenwerte (Scores) der Hauptkomponenten bezogen auf die PCA-Variablen, siehe Gleichung (1.20) und (1.23)

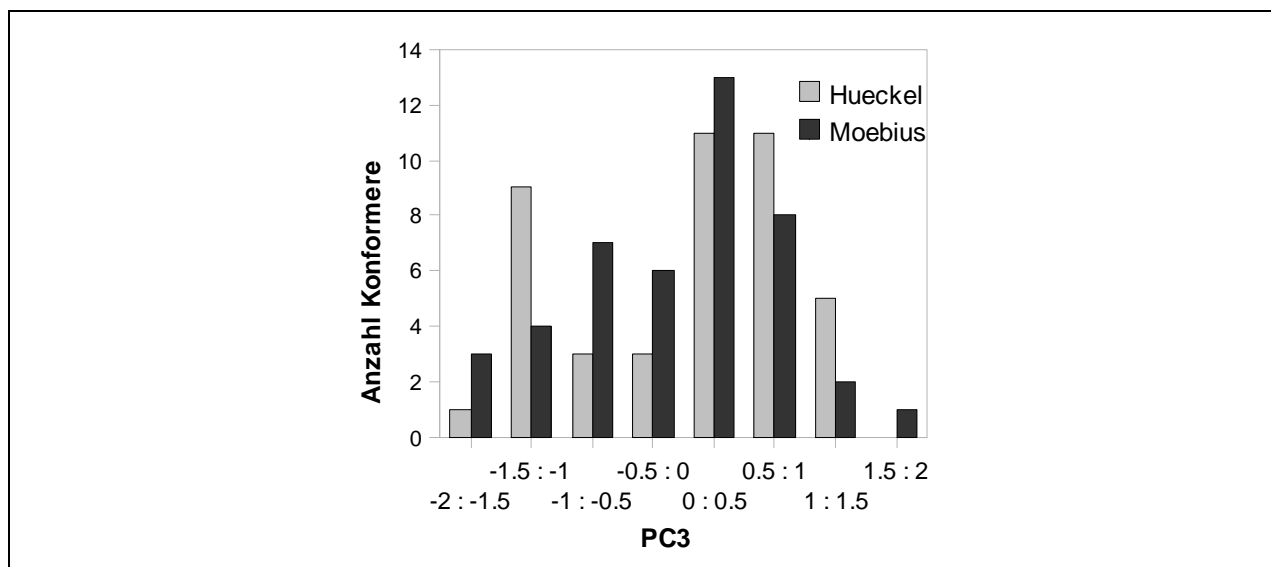
Scores ( $\Gamma$ )	PC_1	PC_2	PC_3	PC_4	PC_5	PC_6	PC_7	PC_8
Tmax	-0.1215	0.7031	0.1034	0.6914	-0.0204	-0.0266	0.0326	-0.0029
HOMA	-0.3455	-0.4243	-0.1979	0.3985	0.0271	-0.0836	0.0494	0.7017
AJ	-0.3456	-0.4245	-0.1973	0.3969	-0.0402	-0.0280	-0.0354	-0.7072
FISE	-0.2236	-0.2207	0.9480	0.0446	-0.0036	0.0220	-0.0100	0.0020
NICS	0.4127	-0.1793	0.0211	0.2966	0.1213	0.8039	-0.2160	0.0387
MAGS	0.4180	-0.1489	0.0690	0.1998	0.6607	-0.2900	0.4823	-0.0722
FISENICS	0.4167	-0.1629	0.0558	0.1838	-0.7360	-0.0844	0.4627	0.0038
FISEMAGS	0.4216	-0.1259	0.0644	0.2048	-0.0648	-0.5036	-0.7084	0.0256

Vereinfacht kann man die ersten beiden Hauptkomponenten als magnetische Kriterien (Hauptachse 1, die wichtigste Achse) und strukturelle Kriterien (Hauptachse 2) betrachten. Die maximale Auslenkung aus der Planarität,  $T_{\max}$ , gehört wie erwartet eher zu den strukturellen Kriterien. Um >95% der Gesamtvarianz der Daten zu erklären, wird die dritte Hauptkomponente benötigt, welche durch den Einfluss der FISE-Energie definiert wird

Zusammengefasst bestätigen die Daten der PCA die bereits verwendete Einteilung der Aromatizitätskriterien in magnetische, strukturelle und energetische Kriterien, wobei die erstere in der Hauptkomponente 1 (PC1) den wichtigsten Einfluss haben. Da 67%, also rund zwei Drittel der Gesamtvarianz, auf die Hauptkomponente 1 entfallen, sind die magnetischen Kriterien mit Abstand am wichtigsten für die Diskussion der Aromatizität der [4n]Annulene.

Nicht überraschend eignet sich die Hauptkomponente 1 am besten zur Differenzierung der Hückel- und Möbiusisomere. Die Histogramme der Verteilung der Hauptkomponenten 1 bis 3 sind in Abbildung 200 dargestellt





**Abbildung 200:** [16]Annulen – Histogramme der Hauptkomponenten (PC1-PC3)

Die transformierten (siehe Gleichung (1.20)) Daten der Hauptkomponentenanalyse sind in Tabelle 87 wiedergegeben. Die laufende Nummer des Isomers in der ursprünglichen Konformationsuche ist in der Spalte „lognr“ angeführt, die Spalten PC\_1, PC\_2, ...PC\_j stehen für die Spaltenvektoren der transformierten Matrix  $x'$  der Eingabedaten (Aromatizitätskriterien).

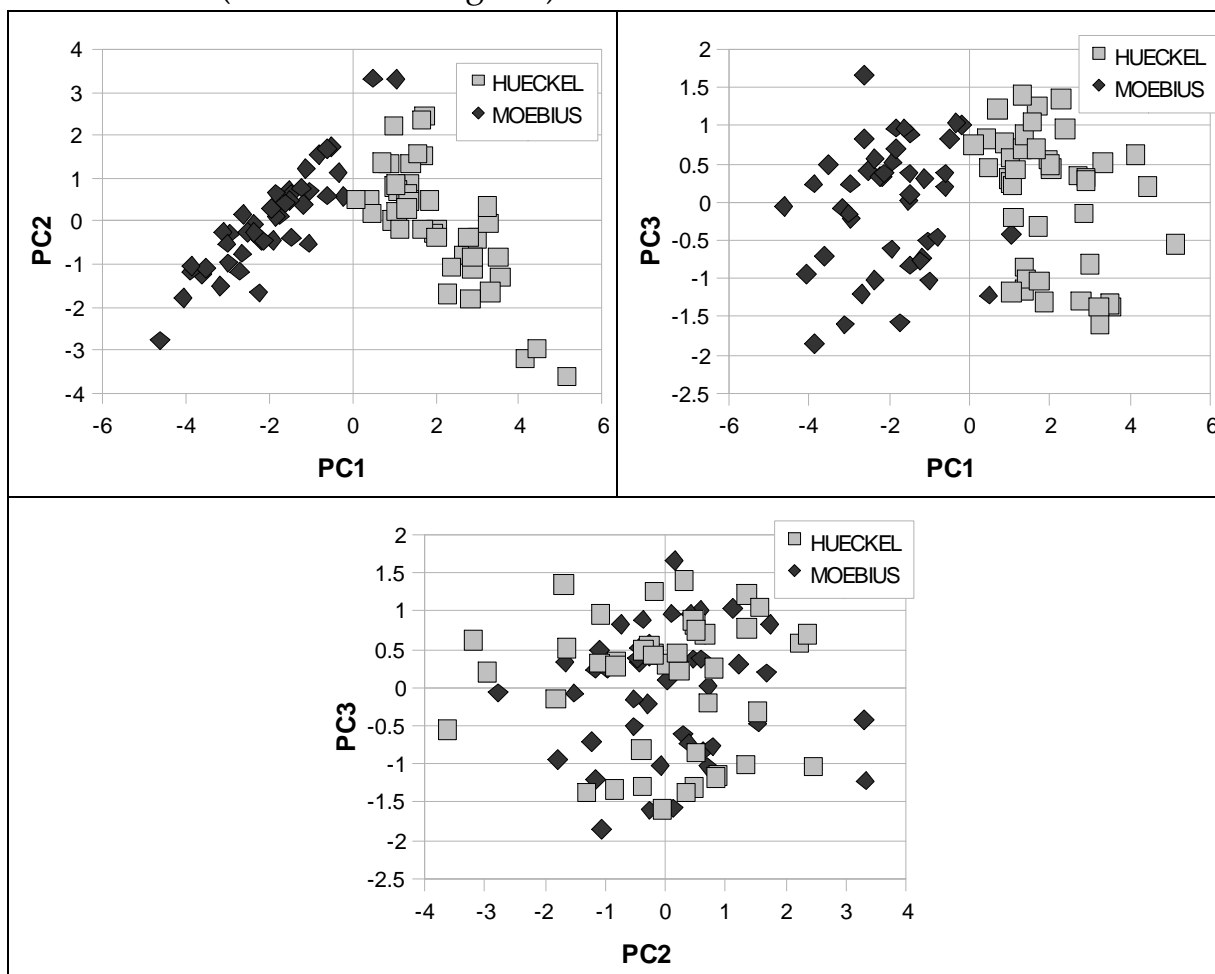
**Tabelle 87:** [16]Annulen – transformierte Matrix  $x'$  der Eingabedaten nach Gleichung (1.20), (*Principal Components Casewise Score*).

lognr	PC_1	PC_2	PC_3	PC_4	PC_5	PC_6	PC_7	PC_8
1	3.54	-1.30	-1.37	0.26	0.12	-0.06	-0.02	0.07
61	3.50	-0.85	-1.34	0.25	-0.02	-0.02	-0.11	0.00
2223	2.02	-0.21	0.44	0.31	0.09	0.05	-0.01	0.06
159	3.31	-1.64	0.52	0.13	0.06	0.07	-0.15	-0.03
260	2.69	-0.81	0.34	0.28	0.04	0.07	-0.04	0.00
9	3.00	-0.41	-0.81	0.51	0.03	-0.06	-0.11	0.05
736	0.87	1.34	0.77	0.34	-0.17	0.06	-0.04	0.01
407	2.88	-1.09	0.32	0.18	-0.03	0.05	-0.02	-0.01
897	0.96	0.02	0.30	-0.22	-0.09	0.00	-0.01	-0.01
791	1.35	0.87	-1.16	-0.64	-0.07	0.00	-0.03	0.03
503	1.94	-0.27	0.55	-0.12	0.01	0.23	0.01	-0.02
143	4.14	-3.20	0.62	0.01	0.21	-0.03	0.03	0.00
621	1.86	0.47	-1.30	-0.18	-0.09	-0.02	-0.04	-0.04
353	3.25	-0.06	-1.60	-0.27	-0.05	-0.06	0.02	0.00
323	1.69	-0.19	1.25	0.24	-0.13	-0.01	-0.04	0.06
2758	1.01	0.81	0.25	-0.95	-0.24	-0.02	0.02	0.04
1007	1.09	0.71	-0.20	-0.62	0.10	-0.07	-0.01	-0.01
254	5.13	-3.62	-0.56	0.10	0.19	0.19	0.05	-0.03
1289	1.07	0.23	0.21	-0.59	-0.22	-0.11	0.01	0.02
222	4.42	-2.96	0.20	0.16	0.13	-0.01	-0.01	0.00
1085	0.70	1.36	1.21	0.43	-0.06	-0.06	0.02	0.03
835	2.88	-0.83	0.28	0.10	-0.01	-0.02	0.04	-0.03
1827	1.69	1.53	-0.32	-0.12	0.05	0.08	0.00	-0.01
480	2.01	-0.37	0.48	-0.16	-0.29	-0.07	0.08	0.03

2625	0.41	0.49	0.82	-0.53	-0.10	-0.10	-0.01	-0.04
1749	0.46	0.19	0.45	-0.90	-0.07	-0.11	0.03	0.00
2116	1.00	2.23	0.58	0.23	-0.01	0.06	0.07	-0.02
1618	1.36	0.51	-0.85	-0.63	-0.09	-0.09	-0.05	0.04
2440	1.39	1.34	-1.01	-0.72	-0.04	-0.04	0.02	-0.01
2191	1.30	0.65	0.70	-0.98	-0.16	0.05	0.01	0.03
1067	1.35	0.46	0.88	0.32	-0.19	-0.15	0.07	-0.01
439	2.27	-1.69	1.35	0.22	0.08	0.13	0.13	0.00
5061	1.56	1.56	1.05	0.30	0.01	0.08	0.01	-0.02
951	2.78	-0.38	-1.30	0.29	-0.19	-0.11	-0.05	0.00
1206	1.12	-0.20	0.43	-0.06	-0.22	-0.23	0.07	-0.04
1720	1.03	0.85	-1.18	-0.88	0.00	-0.03	-0.03	-0.02
586	2.84	-1.82	-0.14	0.02	-0.28	0.26	0.18	0.03
563	1.73	2.45	-1.03	1.30	-0.04	-0.02	0.07	0.01
1692	0.09	0.51	0.75	-1.09	0.02	-0.15	0.07	-0.01
1127	3.23	0.34	-1.37	1.13	0.01	-0.12	-0.03	-0.03
1513	2.38	-1.08	0.96	0.22	-0.04	0.06	-0.10	-0.06
3220	1.30	0.31	1.40	-0.14	-0.04	0.15	0.04	-0.03
2512	1.65	2.36	0.69	-0.83	-0.08	-0.14	0.04	-0.03
1255	-2.19	-0.44	0.32	0.47	0.05	-0.29	0.02	0.02
1548	-2.36	-0.08	-1.02	0.16	0.04	0.06	0.05	0.04
1641	-2.96	-0.29	-0.21	0.79	-0.06	0.07	0.05	0.01
2274	-1.92	-0.44	0.52	-0.35	0.22	0.03	-0.04	0.03
1587	-2.64	-0.75	0.82	0.06	0.18	0.02	0.02	0.03
676	-3.11	-0.26	-1.60	0.32	-0.08	0.14	0.06	-0.01
2162	-2.70	-1.16	-1.20	-0.42	-0.01	-0.04	0.00	0.02
2299	-3.62	-1.23	-0.70	-0.36	-0.22	0.25	-0.03	0.00
2504	-3.89	-1.17	0.24	0.36	0.00	-0.30	-0.07	-0.01
2841	-2.52	-0.27	0.41	0.12	-0.11	0.18	-0.10	-0.01
1498	-1.48	-0.38	0.89	-0.14	0.35	-0.03	0.05	0.02
2059	-4.07	-1.80	-0.94	-0.36	0.13	0.28	0.14	-0.01
4314	-1.53	0.71	0.02	0.17	-0.04	0.07	-0.02	0.00
2764	-1.49	0.62	-0.83	0.11	0.02	0.17	0.01	0.00
1056	-1.05	-0.52	-0.51	0.09	0.28	-0.31	0.10	0.00
1156	-1.74	0.12	-1.58	-0.47	0.50	0.11	0.03	-0.02
3908	-1.83	0.11	0.97	0.24	-0.03	0.09	-0.06	0.04
1875	-4.63	-2.77	-0.06	0.46	-0.42	-0.32	0.03	-0.01
2598	-1.96	0.30	-0.61	0.19	-0.01	0.05	-0.05	0.03
2568	-3.17	-1.51	-0.08	-0.18	0.01	0.10	-0.01	0.00
3240	-1.02	0.70	-1.03	-0.30	0.06	-0.06	-0.01	0.04
2991	-0.21	0.58	1.00	-0.87	0.14	0.09	-0.07	0.02
3170	-1.18	0.38	-0.73	-0.21	0.08	-0.07	-0.03	0.04
3737	-0.50	1.74	0.82	0.46	0.01	0.13	0.00	0.01
3881	-0.82	1.54	-0.46	0.23	0.12	0.05	0.01	0.03
3858	-1.50	0.47	0.37	0.24	-0.03	0.01	-0.05	0.01
3303	-2.64	0.15	1.66	0.25	-0.09	0.07	-0.06	0.03
2321	-2.25	-1.66	0.33	0.31	-0.04	-0.37	0.00	-0.02
3281	0.48	3.32	-1.22	0.78	0.11	-0.07	0.07	-0.02
2586	-0.34	1.12	1.04	0.22	0.22	-0.12	0.08	0.01
4610	-1.12	1.22	0.30	0.58	0.02	0.00	0.01	0.00
5089	-0.62	1.68	0.19	0.39	-0.05	0.20	-0.08	-0.02
3481	-2.98	-0.97	0.24	0.24	0.02	-0.29	-0.01	-0.01

3666	-2.14	-0.46	0.37	-0.28	0.28	0.07	-0.03	0.01
3395	-1.84	0.64	0.70	0.34	-0.01	0.17	-0.03	-0.03
3025	-3.01	-0.52	-0.15	0.15	0.01	0.00	0.02	-0.01
5146	1.04	3.30	-0.42	0.28	0.06	0.05	0.01	-0.02
4414	-1.62	0.41	0.97	0.36	0.04	0.06	-0.02	-0.02
3685	-3.52	-1.10	0.49	-0.13	-0.20	0.27	-0.02	-0.01
3524	-2.36	-0.26	0.56	0.00	0.15	-0.10	0.02	0.00
4794	-0.61	0.59	0.37	-0.55	0.22	-0.21	-0.10	-0.03
3558	-1.23	0.78	-0.77	-0.09	0.05	0.13	0.00	-0.04
4754	-3.88	-1.06	-1.85	-0.21	-0.38	0.09	-0.02	-0.04
4638	-1.49	0.03	0.10	-0.15	0.24	-0.09	-0.03	-0.06

Die Auftragung der Hauptkomponenten gegeneinander bringt keine neuen Erkenntnisse (siehe Abbildung 201).



**Abbildung 201:** [16]Annulen – Auftragung der Hauptkomponenten (PC1-PC3)

Die Trennung der Hückel- und Möbiusisomere entspricht der Trennung, die direkt durch Auftragung der HOMA-Werte über die FISENICS- bzw. FISEMAGS-Werte erhalten werden.

### 3.6.1.8 Vergleich mit Literaturdaten

Die Synthese von [16]Annulen wurde publiziert von Sondheimer<sup>[133]</sup> und später von Oth und Schröder.<sup>[134]</sup> In deutrierter Form wurde [16]Annulen ferner von Stevenson und Kurth untersucht.<sup>[135]</sup> Die Charakterisierung erfolgte

bei tiefen Temperaturen per NMR und im Kristall (Röntgenstruktur), da es bei Raumtemperatur zu intramolekularen Ringschlussreaktionen kommt. Die Struktur eines [16]Annulens in Möbiuskonformation wurde zuerst von Rzepa postuliert.<sup>[124]</sup> Erste, weitgehend unvollständige Versuche, die komplette Hyperfläche des [16]Annulens theoretisch zu beschreiben, finden sich bei Schleyer *et al.*<sup>[125]</sup> Im Rahmen der Synthese des ersten stabilen Möbiusannulens durch Herges *et al.* wurde eine im Rahmen dieser Arbeit angefertigte ausgiebige Isomerenanalyse veröffentlicht.<sup>[127]</sup> Castro, Karney *et al.* beschreiben in ihrer Veröffentlichung von 2006 die dynamischen Prozesse<sup>[136]</sup> bei der Isomerenumwandlung der [16]Annulene auf dem bisher höchsten rechnerischen Niveau und befinden sich dabei in sehr guter Übereinstimmung mit den gemessenen thermodynamischen Literaturwerten.

Die bisher beschriebenen Isomere (B3LYP/6-31G\*-Niveau) sind in Tabelle 88 aufgeführt, zusätzlich einige der besonders „möbiusaromatischen“ neuen Strukturen der Konformationssuche, siehe Tabelle 83. Sämtliche bisher bekannten Strukturen sind enthalten im Datensatz ebendieser Konformationssuche unter der Strukturnummer die in untiger Tabelle als „lognr“ angegeben wird. Damit stellt diese eine Übermenge aller bisher bekannten Strukturen dar.

Einzigste Ausnahme ist Schleiers „Struktur 12“, welche von Schleyer *et al.* als Paradebeispiel für Möbiusaromatizität angeführt wird. Ihre relative Energie liegt mit 51,4 kcal/mol bezogen auf das globale Minimum allerdings dermaßen hoch, dass sie experimentell wohl nicht im Gleichgewicht vorliegt und auch synthetisch nicht zugänglich sein sollte. Aufgrund der Boltzmann-Statistik kann „Struktur 12“ bei Betrachtung der Hyperfläche ignoriert werden.

**Tabelle 88:** [16]Annulen – Zuordnung der Isomere Oths<sup>[137]</sup>, Rzepas<sup>[124]</sup> und Schleiers<sup>[125]</sup> zur KONAN-Analyse aus Tabelle 83

Literaturisomer	CANE	Topologie	Erel[kcal/mol]	NICS	lognr
Rzepas „structure 7“*	0001111100011111	Hückel	8.9	-0.5	2758
Rzepas Moebius Struktur <sup>[124]</sup> *	0001101100011111	Möbius	7.1	-5.2	2274
S <sub>4</sub> global Minimum**	0011001100110011	Hückel	0.0	10.9	1
Oths C <sub>1</sub> Struktur im Gleichgewicht mit dem globalen	0011011011001111	Hückel	2.0	10.7	4

Minimum**					
Konformations- suche 2504 C <sub>2</sub>	0000110110000111	Möbius	7.7	-10.3	2504
Konformations- suche 2059 C <sub>1</sub>	0000101100011011	Möbius	8.5	-8.7	2059
Konformations- suche 1875 C <sub>1</sub>	0000010000110011	Möbius	9.4	-10.8	1875
Konformations- suche 4754 C <sub>1</sub>	0001101011001111	Möbius	16.7	-10.1	4754
Schleyers <sup>[126]</sup> C <sub>2</sub> Struktur 9	0000110011000111	Möbius	5.1	-6.3	1255
Schleyers <sup>[126]</sup> C <sub>2</sub> Struktur 10	0001100011000111	Möbius	7.6	-3.9	1587
Schleyers <sup>[126]</sup> C <sub>2</sub> Struktur 11	0000010000011011	Möbius	15.8	-14.5	5295
Schleyers <sup>[126]</sup> C <sub>2</sub> Struktur 12	0000011111111111	Möbius	51.4	-15.6	-

\* Im *supporting material*<sup>[124]</sup> wird von Rzepa eine Struktur angegeben, welche mit einem CANE-Deskriptor von 0001111100011111 (entsprechend lognummer=2758) nicht mit der im Artikel besprochenen Konfiguration und Topologie übereinstimmt. Die abgebildete Struktur in Fig. 1 lässt die Ermittlung des CANE-Wertes zu, welcher sich von der Struktur im *supplementary material* unterscheidet. Ebenso stimmt die Topologie der Koordinaten (Hückel) nicht mit der besprochenen (Möbius) überein. Es wird hier der beste Kandidat aus Tabelle 83 zugeordnet.

\*\*Das S<sub>4</sub>-Isomer des [16]Annulen wurde von Oth synthetisiert und durch Röntgenstrukturanalyse charakterisiert. Dabei wurden geringen Mengen der C<sub>1</sub>-Struktur im Gleichgewicht mit der stabileren S<sub>4</sub>-Struktur gefunden.<sup>[133],[134],[137]</sup>

Schleyer *et al.* beanspruchen in ihrer Veröffentlichung,<sup>[125]</sup> die ersten echten Möbius-[16]Annulen-Isomere gefunden zu haben, da sie Rzepas Struktur mit einem NICS-Wert von -0.5 als nichtaromatisch klassifizieren. Dies dürfte ein Irrtum sein, beruhend auf der Verwendung der offensichtlich fehlerhaften Koordinaten aus Rzepas „structure 7“, siehe Anmerkungen zu Tabelle 87. Die tatsächliche Struktur weist mit -5.2 einen mittelstarken magnetisch aromatischen Charakter auf. Die „Ehre der ersten Struktur“ gebührt damit wohl zweifelsfrei Rzepa *et al.*



Die Abbildung 202 zeigt die Einordnung der publizierten Isomere auf der vollständigen Hyperfläche der KONAN-Analyse. Aufgetragen werden die NICS-Werte über die HOMA-Werte.

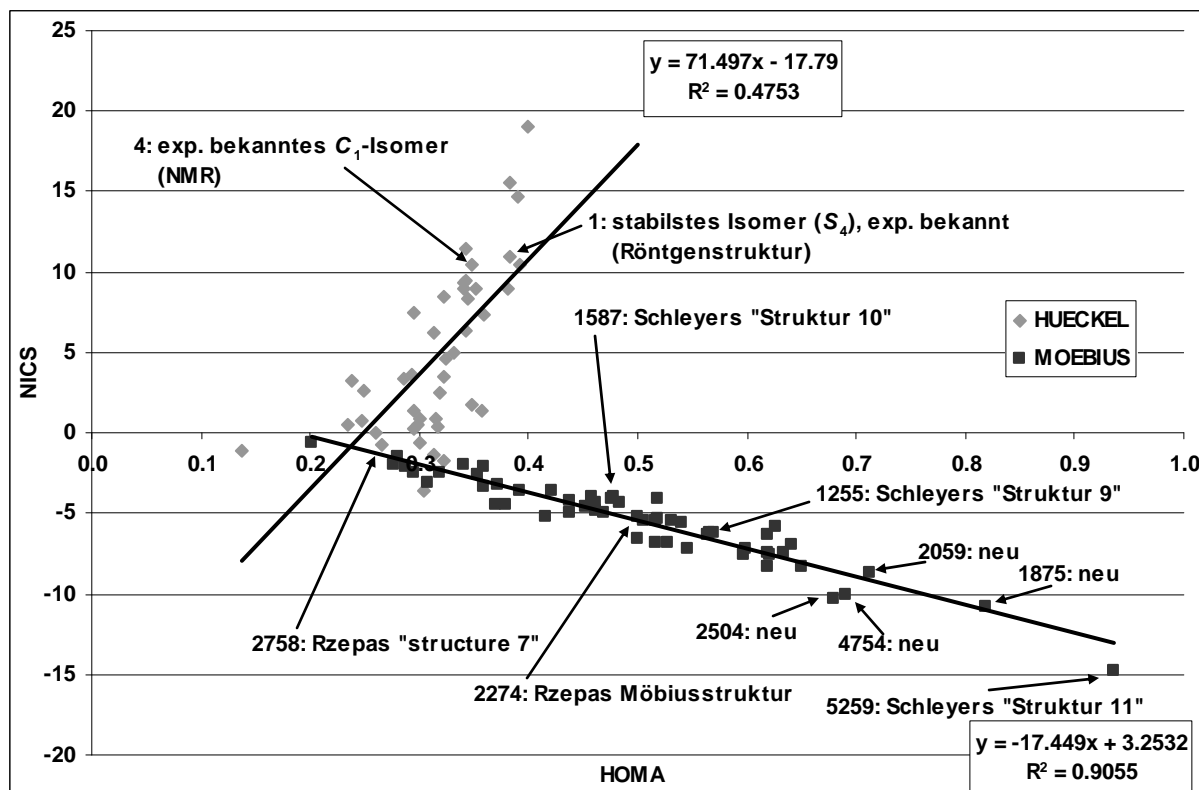


Abbildung 202: [16]Annulenes – Einordnung literaturbekannter Strukturen in die KONAN-Analyse aus Tabelle 83, vor dem Doppelpunkt angegeben die „lognr“ der ursprünglichen Konformationsuche

### 3.6.2 Andere Kraftfelder

Der Versuch, verwendbare Startstrukturen mit Hilfe des Kraftfeldes OPLS und der *conformational search* Routine des Programms HYPERCHEM zu erhalten schlägt leider fehl, da die Strukturen während der B3LYP/6-31G\*-Optimierung starke Konformations- und Konfigurationsänderungen erfahren. Dies stellt die Kongruenz der DFT-Hyperfläche zur OPLS-Hyperfläche in Frage, die erhaltenen Daten werden daher verworfen.

Die Verwendung des Programms *Sybyl* 6.9<sup>[39]</sup> und des Konformationsuchprogramms DISCO ergibt eine vergleichbare Zahl Isomere (395). Die Optimierung dieser Strukturen ergibt eine Hyperfläche, die weitgehend (die ersten 5 Strukturen sind exakt identisch) mit der MMP-Hyperfläche übereinstimmt. Auf weitere Rechnungen wird deshalb verzichtet. Die tabellierten Ergebnisse sind auf der beigelegten DVD zu finden.

### 3.7 Das erste Möbiusannulen

Die erste stabile (und durch Röntgenstruktur belegte) Möbiusstruktur wurde, wie in Kapitel 3.1 beschrieben, von Herges und Ajami publiziert.<sup>[84]</sup> Dabei handelt es sich formal um ein bianthrachinodimethan-modifiziertes [16]Annulen, siehe Abbildung 171.

Schleyer *et al.* bezweifeln in ihrer Veröffentlichung<sup>[126]</sup> von 2005 die Aromatizität dieser Verbindung mit dem nicht unberechtigten Argument, dass die zum Teil recht hohen Abweichungen aus der Planarität die für die Aromatizität wichtige Überlappung der  $\pi$ -Orbitale stark verringern.

Im Vergleich mit klassischen Aromaten wie Benzol sind die beobachteten aromatischen Effekte gemessen an strukturellen, energetischen und magnetischen Kriterien tatsächlich relativ gering.

Die Wahrheit in dieser Kontroverse liegt wahrscheinlich zwischen den Extrempositionen „aromatisch“ und „nichtaromatisch“, wie mit den folgenden Daten belegt werden soll, welche auch Teil unserer publizierten Erwiderung sind.<sup>[127]</sup>

Die Bianthrachinodimethan-Einheit wurde, wie bereits in der Einleitung dieses Kapitels erwähnt in das [16]Annulen eingeführt, um die Verdrillung zur Möbius-Topologie zu stabilisieren. Drei Doppelbindungen und vier Einfachbindungen sind damit in ihrer E,Z-Konfiguration, bzw. ihrer s-cis-/s-trans-Konformation festgelegt und es verbleiben 72 mögliche CANE-Isomere, deren Aromatizität im Folgenden mit den bereits erwähnten Methoden systematisch untersucht wird. Es sind verschiedene [16]Annulen-Pfade cyclischer Konjugation denkbar. Aufgeführt sind diese in Abbildung 203. Wie in der Publikation von Schleyer *et. al.* wird das arithmetische Mittel der mit ihrer Hilfe erhaltenen Aromatizitätskriterien verwendet, siehe Tabelle 89.

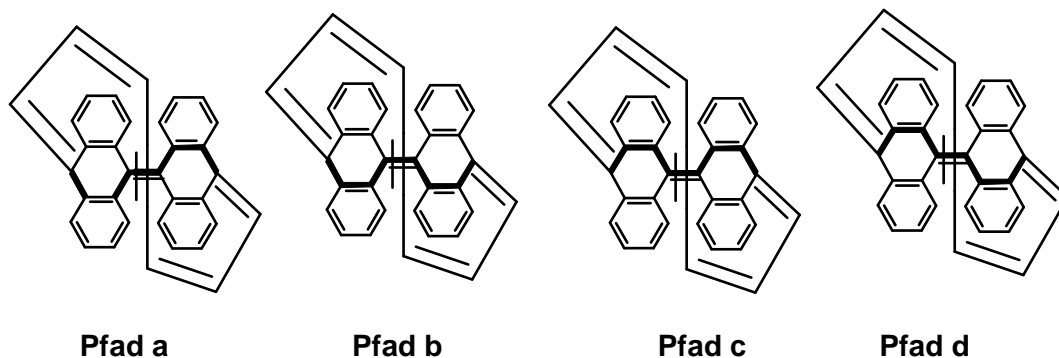


Abbildung 203: Bianthrachinodimethan-modifizierte [16]Annulene

**Tabelle 89: Bianthrachinodimethan-modifiziertes [16]Annulen, KONAN-Analyse, arithmetisches Mittel aller Pfade, Energien in [kcal/mol]**

lognr	Erel	Topology	Tmax*	HOMA	AJ	NICS	MAGS	ISE
6, globales Minimum	0	MOEBIUS	70.83665	0.02555	0.70925	-2.6259	-230.915	
25	0.22	MOEBIUS	66.72385	0.085625	0.73165	-3.2795	-239.352	
9	0.61	MOEBIUS	64.7207	-0.01465	0.69645	-1.8026	-232.202	
7	0.63	MOEBIUS	65.2052	0.030425	0.71305	-1.8418	-232.723	
17	2.39	MOEBIUS	62.78045	0.17145	0.77	-4.0408	-252.819	
21, C <sub>1</sub> X-ray	2.77	MOEBIUS	64.63945	-0.09965	0.672475	-0.9181	-230.026	
23, C <sub>2</sub> X-ray	2.87	MOEBIUS	71.09833	-0.00325	0.700325	-2.8269	-229.067	4.04
4	3.66	HUECKEL	68.54295	-0.13945	0.647675	-1.2906	-228.722	
19	6.82	MOEBIUS	68.97475	-0.11595	0.662175	-1.0593	-226.682	
12, C <sub>s</sub> X-ray	6.98	HUECKEL	76.9411	-0.267975	0.60425	-1.2556	-228.916	-2.22
16	8.33	MOEBIUS	60.6484	0.13005	0.7579	-3.8976	-251.019	
8	9.30	MOEBIUS	75.46775	-0.039325	0.689525	-2.1031	-229.719	
18	9.34	MOEBIUS	53.68913	0.23655	0.804725	-6.1973	-271.398	
24	11.10	MOEBIUS	73.0224	-0.095225	0.66895	-1.673	-228.158	
3	11.86	HUECKEL	59.67168	-0.1182	0.671125	0.9669	-210.715	
10	12.44	MOEBIUS	71.4161	-0.1978	0.63135	-1.3392	-226.868	
11	12.63	MOEBIUS	70.12245	-0.159375	0.6433	-1.4067	-226.917	
2	14.33	MOEBIUS	67.14175	-0.10365	0.6783	-1.6405	-234.823	
5	16.08	MOEBIUS	79.322	-0.0114	0.706	-2.44	-229.352	
15	23.50	MOEBIUS	65.6747	-0.061625	0.69375	-2.5984	-233.861	
14	23.73	HUECKEL	60.7902	0.032525	0.7246	1.0115	-205.456	
1	27.09	HUECKEL	62.6555	0.1469	0.77135	1.5096	-199.785	
13	28.40	HUECKEL	60.00855	0.09595	0.753675	1.8463	-200.836	
20	30.36	MOEBIUS	72.4221	0.12825	0.758475	-4.0261	-241.01	

\*: angegeben wird das Minimum über alle Pfade

\*\* : keine *cis*oide Doppelbindung für die ISE<sub>II</sub>-Derivatisierung im Molekül

ISE<sub>II</sub>-Energien wurden bisher nur für einzelne Strukturen erhalten und bestätigen die Erwartungen einer schwachen Möbiusaromatizität bzw. Hückelanti-aromatizität.

Eine PCA über die Einflussgrößen Tmax, HOMA, Aj, MAGS und NICS zeigt, dass 87% der Varianz bereits durch die beiden Hauptachsen erklärt werden,

**Tabelle 90:** PCA der bianthrachinodimethan-modifiziertes [16]Annulen-Isomere - Hauptkomponenten, ihre Eigenwerte und Anteile an der Gesamtvarianz

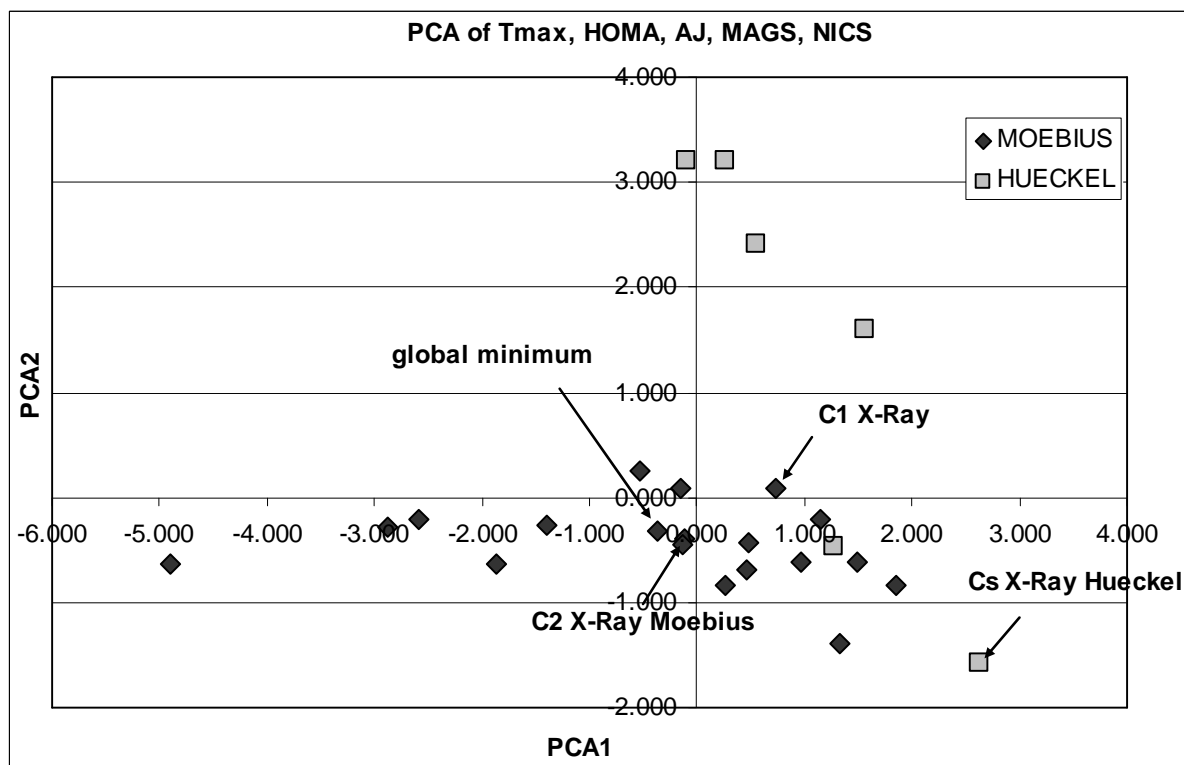
	PC_1	PC_2	PC_3	PC_4	PC_5
Eigenwert	2.786	1.586	0.599	0.023	0.006
%Varianz	55.715	31.725	11.980	0.461	0.120
%Varianz, kumulativ	55.715	87.439	99.419	99.880	100.000

Hauptachse 1 repräsentiert im wesentlichen die strukturellen Kriterien, Hauptachse 2 die magnetischen Kriterien, wie aus Tabelle 91 zu erkennen ist.

**Tabelle 91:** PCA der bianthrachinodimethan-modifiziertes [16]Annulen-Isomere - Eigenwerte (Scores) der Hauptkomponenten bezogen auf die PCA-Variablen, siehe Gleichung (1.20) und (1.23)

Scores (I)	PC_1	PC_2	PC_3	PC_4	PC_5
Tmax	0.303	-0.465	0.817	0.147	-0.044
HOMA	-0.534	0.287	0.349	-0.145	-0.700
AJ	-0.521	0.317	0.363	0.261	0.654
MAGS	0.417	0.538	0.279	-0.654	0.177
NICS	0.422	0.558	0.027	0.680	-0.220

Die Auftragung der beiden Hauptachsen ergibt dabei eine klare Unterscheidung in aromatische, nichtaromatische und antiaromatische Bereiche, siehe Abbildung 204.



**Abbildung 204:** Bianthrachinodimethan-modifizierte [16]Annulen-Isomere - aufgetragen sind die Hauptachsen der PCA, wobei ausgewählte Strukturen samt ihrer „lognr“ aus der Konformationsuche zugeordnet werden.

Die Hückelröntgenstruktur liegt tief im nichtaromatischen Bereich, während zumindest die C<sub>2</sub>-Röntgenstruktur relativ solide im aromatischen Bereich eingeordnet werden kann. Anhand einzelner Aromatizitätskriterien ist die Klassifizierung der Aromatizität für die Isomere des bianthrachinodimethan-modifizierten [16]Annulens nicht zweifelsfrei möglich, was Schleyer *et al.* zum Schluss kommen lässt, dass die Verbindung aufgrund der Benzanellierung nichtaromatisch ist. Eine Untersuchung mittels PCA kann aber trotz der Überlagerung durch sterische und elektronische Einflüsse im Molekül (die anellierten Benzolringe) durchaus einen Einfluss der aromatischen Konjugation belegen.

Es kann hier vom ersten Möbiusaromaten gesprochen werden, auch wenn es sich im Vergleich zu einigen theoretischen Isomeren des [16]Annulen-Stammsystems oder gar dem Benzol eher um schwach ausgeprägte Aromaten handelt. Zum Abschluss wird in Abbildung 205 das C<sub>2</sub>-symmetrische aromatischste Isomer mit eingezeichneten Trajektorien entlang der  $\pi$ -Delokalisierung eines [16]Annulen-Pfades dargestellt.

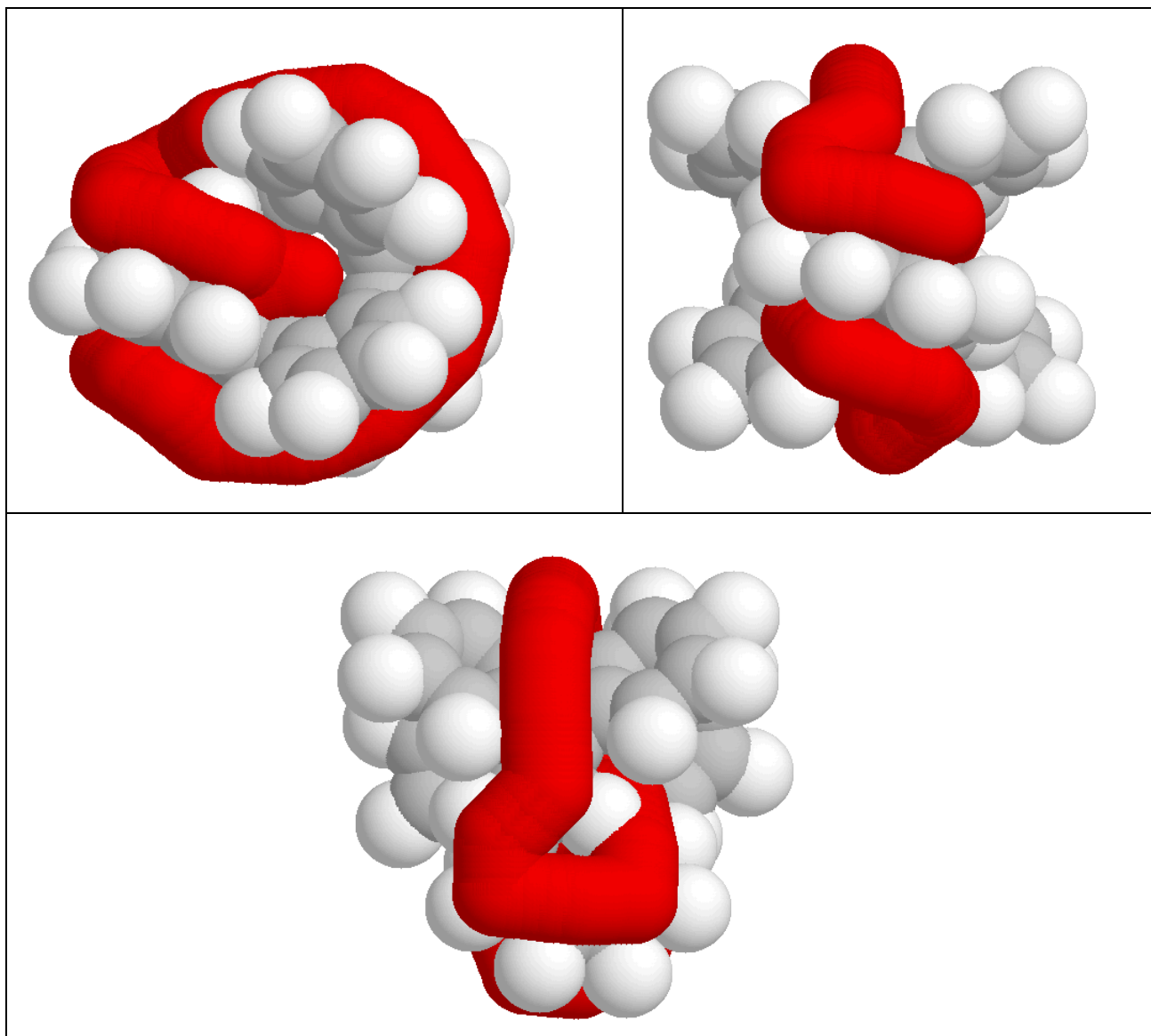


Abbildung 205: Bianthrachinodimethan-modifiziertes  $[16]$ Annulen – rote Trajektorien entlang der p-Orbitale zeigen die Möbiustopologie der  $C_2$ -Röntgenstruktur<sup>[84]</sup>

## 4 Zusammenfassung und Ausblick

### 4.1 Kohonennetze zur Konformationsanalyse (NEURO)

Im Rahmen dieser Arbeit werden Kohonennetze erfolgreich zur Projektion hochdimensionaler Eingabedaten auf eine zweidimensionale Karte verwendet. Als automatische *unsupervised-learning*-Methode sind diese Netzkarten sehr attraktiv, um komplexe und unübersichtliche Hyperflächen anschaulich zu machen. Sie stellen ein gutes, neues Werkzeug dar, die Konformationsräume aus systematischen Konformationsanalysen, Moleküldynamiksimulationen oder Monte-Carlo-Simulationen auszuwerten und zu beurteilen. Insbesondere das TRAJECTORIX-Modul hat sich dabei im Rahmen dieser Arbeit bewährt.

Die in dieser Arbeit entwickelten Methoden und Programme sind gebündelt unter dem Namen „NEURO-Suite“ unter anderem auf der beigelegten DVD oder direkt vom Author erhältlich. Anwendungen auf interessante neue Substanzen sind Gegenstand von Kooperationen innerhalb der Arbeitsgruppe und mit anderen Arbeitsgruppen (z.B. das Hexabenzocoronen in Kooperation mit Dötz,<sup>[120]</sup> Kapitel 2.8.9, diverse Zuckerdendrimere mit Fuchs<sup>[138]</sup>, hier aus Platzgründen nicht ausgeführt, Untersuchungen der Konformationsflexibilität an Liganden für Anionen<sup>[139]</sup>, ebenfalls nicht angeführt).

#### 4.1.1 Standardtrainingsparameter

Die Wahl der Netztrainingsparameter ist eine Optimierungsaufgabe, für die es kein erkennbares allgemeingültiges Optimum gibt. Für bekannte Hyperflächen kleiner Moleküle werden sie im Rahmen dieser Arbeit systematisch variiert. Aufbauend darauf werden folgende robuste NEURO-Standardtrainingsparameter vorgeschlagen.

- epochs: in Anbetracht der erhaltenen MTS-Werte (siehe Gleichung (2.26)) wird folgende Gleichung zur Bestimmung der Trainingsdurchläufe vorgeschlagen.

$$(3.1) \quad \text{epochs} = \frac{10^7}{\mu_a(\text{Anzahl der Strukturen im Trainingssatz})}$$

- NETX, NETY (Zahl der Neuronen): je nach Umfang der Hyperfläche. 25x25 als Vorgabe.
- $\Delta_{\max}(1)$ : Kann aus der Netzgröße errechnet werden:

$$(3.2) \quad \Delta_{\max}(1) = 0.4 \max(\text{NETX}, \text{NETY})$$

- $f_0=0.15$
- $\eta(\xi)=\text{linear}$
- Abstandsfunktion: Von den untersuchten Netzfunktionen konnten diejenigen mit negativen Anteilen (*Mexican hat* und *pseudo Mexican*) als unbrauchbar für die vorliegende Anwendung identifiziert werden. Die übrigen teilen sich auf in drei Gruppen, die je nach untersuchter Hyperfläche Vor- und Nachteile haben. Als Standardfunktional hat sich das „parabolische“ bewährt.

**Tabelle 92: Netzfunktionale – Übersicht der Charakteristika**

Gruppe	Funktionale	Charakteristika
A	<i>block</i>	starke Vereinfachung der Netzkarte, geringer Detailreichtum
B	<b>parabolisch</b>	ausgewogenes Verhältnis von Vereinfachung und Detailreichtum
C	triangular quadratisch-parabolisch quadratisch-triangular	komplexe Netzkarten, hoher Detailreichtum

#### 4.1.2 Vergleich mit konkurrierenden Verfahren

Die direkte Konkurrenz der Kohonennetze dürfte in deterministischen Verfahren, wie z.B. dem *Multidimensional Scaling*<sup>[50],[53]</sup> bestehen, welches ebenfalls eine Projektion von Hyperflächen und Trajektorien ermöglicht. Auch Clusterverfahren werden oft zu Konformationsanalysen herangezogen. Die Vorteile und Nachteile der Kohonennetze werden im Folgenden besprochen.

##### 4.1.2.1 Stärken

###### Rückkopplung der Parameter:

Der große Vorteil der Kohonennetze liegt in der Möglichkeit, die Netzparameter auf die Prototypstrukturen im Rahmen des MINIX-Moduls rückkoppeln zu lassen. Damit werden für verrauschte Daten, z.B. aus MD-Rechnungen, dramatische Verbesserungen in der Qualität der erhaltenen Netzkarten erzielt. In der Tat sind MD-Netzkarten, die ohne Optimierungsschritt aus dem COLORIX-Modul erhalten werden, in der Regel unbrauchbar im Vergleich zu den optimierten Netzkarten. Lediglich für voroptimierte Strukturen, z.B. erhalten beim Hexabenzocoronen, kann auf den Optimierungsschritt verzichtet werden. Beim *Multidimensional Scaling* ist eine solche Rückkopplung prinzipbedingt nicht möglich.



### Autoclustering

Dieser Punkt ergibt sich aus dem zweiten Optimierungsschritt, MINIX2. Eine Auftragung der Gradienten unterteilt die Netzkarte visuell in Konformationsbereiche, womit die Trajektorie übersichtlich zugeordnet werden kann. Das *Multidimensional scaling* ergibt in der Regel zu geringe Differenzen der Gradienten für ein erfolgreiches *Autoclustering*.

### Übergangszustände

Auch Übergangszustände zwischen zwei Konformationsbereichen werden nach Optimierung bei angewandten *restraints* (MINIX1) meist gut repräsentiert. Dies kann die Ermittlung der stationären Punkte auf einer zunächst unbekanntem Hyperfläche stark vereinfachen.

### Rechenzeit

Da nur die Prototypstrukturen minimiert werden, kommt es zu einer signifikanten Verringerung der Anzahl der zu optimierenden Strukturen, gerade wenn in langen MD-Trajektorien in den sonst häufig angewandten Konformationsverfahren in regelmäßigen Abständen Strukturen extrahiert und minimiert werden. Dies wirkt sich positiv auf die Rechenzeit aus. Die Projektion von Strukturen auf die Hyperfläche nach erfolgtem Training des Netzes erfordert praktisch keinen Rechenaufwand mehr.

### Gezielte Aufbereitung der Daten

Isomerendatensätze können durch Wahl zu trainierenden Eingangsparameter gezielt in Hinblick auf bestimmte Eigenschaften wie Abstände zweier reaktiver Gruppen untersucht werden.

#### **4.1.2.2 Schwächen**

Auch die Nachteile sollen nicht verschwiegen werden:

### Begrenzte Komplexität

Hyperflächen mit vielen Freiheitsgraden und Minima können prinzipbedingt nicht mehr übersichtlich durch Kohonennetze dargestellt werden und enthalten viele Projektionsfehler. Das untersuchte n-Oktan ist in dieser Beziehung bereits ein relativ schwieriger Kandidat. Glücklicherweise sind biologisch relevante Konformationshyperflächen in der Regel sehr eingeschränkt, so dass dies meist keinen Nachteil darstellen sollte. Ein Beispiel für völlig ungeeignete

Moleküle sind z.B. langkettige Alkane. Ein Beispiel für gut geeignete Moleküle sind Hormone, für die das *cis-Decalin* als Modell gewählt wurde.

Die begrenzte Komplexität der Eingabedaten teilen die Kohonenkarten mit der direkten Konkurrenz, dem *Multidimensional Scaling*,<sup>[50]</sup> im Gegensatz zu den *Cluster-Analysis*-Verfahren.

#### Manuelle Wahl der Eingabekoordinaten

Die Wahl der zu trainierenden internen Koordinaten eines Systems erfolgt im Augenblick noch manuell und nach chemischem Wissen. Es muss im Voraus abgeschätzt werden, welche internen Koordinaten eines Systems von Bedeutung sein können. Dies verhindert automatisierte Lösungen und erschwert die Benutzung des Programms. Außerdem sind Netzkarten, die mit unterschiedlichen Koordinaten erzeugt werden, nur bedingt vergleichbar.

#### **4.1.3 Ausblick**

Weitere Anwendungen und Verbesserungen der Kohonennetze zur Konformationsanalyse werden nachfolgend vorgeschlagen.

#### Anwendung auf größere Biomoleküle

Die Anwendung auf größere Biomoleküle scheiterte leider an der begrenzten Leistungsfähigkeit des Programms HYPERCHEM bei der Generierung und Optimierung der Eingabedaten. Vorversuche im Rahmen dieser Arbeit deuten darauf hin, dass die Verwendung von *NMR-restraints* bei der Erstellung der Eingabedaten einen wichtigen Schritt bei der Betrachtung der Hyperfläche bioaktiver Substanzen sein sollte.

#### Entkopplung der NEURO-Suite von HYPERCHEM

Im Augenblick besteht eine große Abhängigkeit der NEURO-Suite vom proprietären Programm HYPERCHEM. Dadurch ermöglicht wurde in der Prototypphase ein einfaches Experimentieren mit der Vielzahl der in HYPERCHEM implementierten Kraftfelder und Methoden sowie eine effiziente Verwendung der leistungsfähigen Schnittstelle sowie der gut dokumentierten Binärformate.

In Hinblick auf die Portierbarkeit und die recht begrenzte *Performance* des Programms HYPERCHEM ist eine Entkopplung und Verwendung von „*Open Source*“ Programmen für MM-Rechnungen wünschenswert.

### Automatische Wahl der Trainingskoordinaten

Die Wahl der Eingabedaten sollte sich prinzipiell mit Hilfe einer vorgeschalteten PCA der Eingangsdaten (z.B. der Distanzmatrix) automatisieren lassen. Der einfachste Ansatz besteht im Training der internen Koordinaten welche als Komponenten die höchste Beladung der Hauptkomponenten in der PCA erzielen. Eine komplexere, aber wahrscheinlich korrektere Implementierung würde die Hauptkomponenten direkt trainieren und auch als *restraints* verwenden, was eine Anpassung des Kraftfeldes in den MINIX-Modulen erfordert. Im Augenblick scheitert dies an der engen Kopplung an das Programm HYPERCHEM, für welches kein Quelltext vorliegt.

### Kombination mit Clusteringverfahren

Interessant wäre es sich auch, die Prototypstrukturen einer Netzkarte einem Clusteringverfahren zuzuführen und die erhaltenen Cluster auf der Netzkarte zu visualisieren.

### Prototypensembles für Neuronen

Im Augenblick wird jedes Neuron durch genau eine Prototypstruktur repräsentiert. Denkbar ist aber auch, dass dies durch eine nach Ähnlichkeit sortierte und gewichtete Liste an Isomeren geschieht. Die gewichtete Mittelung von Moleküleigenschaften und eventuell der Struktur über dieses Ensemble ermöglicht unter Umständen einen Verzicht auf den ersten Optimierungsschritt. Dies würde eine signifikante Beschleunigung des Verfahrens sowie eine vereinfachte Verwendung externer Modellierungsprogramme bedeuten.

## **4.2 Aromatizität von Annulen-Isomeren**

### **4.2.1 Zusammenfassung**

Die KONAN-Analyse wird im Rahmen dieser Arbeit als Methode entwickelt, schnell und effizient die stabilsten Isomere von Annulenen finden und auf Aromatizität zu untersuchen. Anwendungen auf das [12]Annulen und das [16]Annulen ergeben die ersten vollständigen Hyperflächen auf B3LYP/6-31G\*-Niveau. Die übereinstimmenden Ergebnisse der Konformationssuchen MMP/HYPERCHEM sowie MMFF94/DISCO/Sybyl lassen diesen Schluss zu. Ermöglicht wird dies durch die Kopplung der verschiedenen *Modelling*-Methoden (Kraftfeld, Semiempirik, DFT) im Rahmen der KONAN-Methode. Es muss darauf hingewiesen werden, dass die Hyperflächen auf einem anderen Niveau quantitativ und qualitativ zum Teil unterschiedlich aussehen

können, was Rechnungen mit anderen Funktionalen (KMLYP sowie BHandHLYP) ergeben haben. Inzwischen existieren zu beiden Annulenen Rechnungen auf höherem Niveau, die aber zumindest teilweise auf den Ergebnissen der Konformationssuche dieser Arbeit aufbauen ([12]Annulen,<sup>[140]</sup> [16]Annulen<sup>[136]</sup>)

Eine Voraussetzung für die effiziente automatische Verarbeitung der großen Mengen an Isomeren liegt dabei in der Erkennung der identischen Isomere durch die ebenfalls neu eingeführten CANE-Deskriptoren, einem speziellen Nomenklaturschema für Annulene, implementiert im Programm XYZ2CANE.EXE.

Ein weiteres Ergebnis dieser Arbeit ist eine Variation der ISE<sub>II</sub>-Methodik<sup>[66]</sup> (FISE), welche drei neue Aromatizitätskriterien ergibt:

- FISE (energetisch),
- FISENICS (magnetisch) und
- FISEMAGS (magnetisch).

Diese sind ebenso wie die bekannten ISE-Werte unter der Voraussetzung erhältlich, dass mindestens eine substituierbare *cis*-Doppelbindung im zu untersuchenden Ring vorliegt. Der Vorteil gegenüber den ISE-Werten liegt gerade für gespannte Strukturen in der schnelleren Konvergenz der benötigten Optimierungen und in der Konfigurationsstabilität des Ringes. Wie in Kapitel 3.6.1.7 ausgeführt, erweisen sich die FISEMAGS-Werte als die besten magnetischen Kriterien zur Trennung von aromatischen und nicht aromatischen Strukturen. Der Wert des Vorzeichens ist außerdem hinreichend (negative Werte sind aromatisch, positive sind antiaromatisch), um Isomere referenzfrei zu klassifizieren. Dies verschafft den FISE-Deskriptoren eine Sonderrolle gegenüber den einfachen Deskriptoren wie NICS, MAGS oder HOMA.

Als Fazit für die Aromatizität in [4n]Annulenen bleibt die Aussage, dass es drei verschiedene Arten von Aromatizitätskriterien gibt: magnetische, strukturelle und energetische. Insbesondere die ersten beiden Kriterien verhalten sich für Hückelisomere der [4n]Annulene unterschiedlich zu den Möbiusisomeren.

Für Möbiusisomere (aromatisch) sind die strukturellen Kriterien positiv mit den magnetischen korreliert, für Hückelisomere (antiaromatisch) ist die Korrelation negativ.

Strukturelle und energetische Kriterien werden durch „Störungen“ struktureller Art (Verdrillung und sterische Hinderung) und elektronischer Art (Benzanellierung) relativ stark beeinflusst. Der strukturelle und energetische aromatische Charakter geht so auch beim bianthrachinodimethan-modifizierten [16]Annulen weitgehend verloren.

Trotzdem konnte die magnetische Aromatizität des bianthrachinodimethan-modifizierten [16]Annulens in Kapitel 3.7 eindeutig belegt werden, so dass mit Recht vom ersten Möbiusaromaten gesprochen werden darf.

#### **4.2.2 Ausblick**

Rechnungen auf höherem Niveau sind wie beschrieben bereits durchgeführt und publiziert<sup>[84],[127]</sup> für die [12]- und [16]Annulene.

Besonders interessant in Hinblick auf Synthese und Stabilität scheinen die kationischen Möbiusannulene  $[9^+]$ ,  $[13^+]$ ,  $[17^+]$  zu sein. Die rechnerische Untersuchung der Hyperfläche und die Synthese und Charakterisierung dieser Verbindungsklassen sind Themen weiterer Arbeiten der Arbeitsgruppe Herges.

## 5 Appendix A: Verwendete Programme und Hilfsmittel

Es folgt eine Auflistung der selbst geschriebenen Programme sowie der verwendeten externen Hilfsmittel. Im Interesse der Zugänglichkeit für einen möglichst großen Teil der wissenschaftlichen Gemeinschaft sind die Hilfetexte und die interne Dokumentation der selbst geschriebenen Programme in englischer Sprache verfasst. Teile des folgenden Kapitels sind als Auszüge aus den Programmquellen realisiert und liegen ebenfalls auf Englisch vor.

### 5.1 Externe Hilfsmittel

Sämtliche eigenen Programme sind in den Programmiersprachen C/C++ (mit externer Verlinkung an C-Bibliotheken von Fremdanbietern) sowie der Skriptsprache Tcl/Tk verfasst.

#### 5.1.1 Modellingprogramme

Als externes Modellingprogramm kommt in dieser Arbeit vor allem das Programm HYPERCHEM<sup>[37],[43]</sup> zum Einsatz. Von Vorteil erweist sich dabei die Integrationsfähigkeit mit eigener Software. Die komplette Grundfunktionalität des Programms ist mittels des *HYPERCHEM Application Programming Interface* (HAPI) (hapi.dll für C/C++, thapi.dll unter Tcl/Tk) per DDE ansteuerbar. Dabei agiert HYPERCHEM quasi als *Compute-Server* und realisiert den Zugriff auf mehrere unterschiedliche Rechenmethoden, die von verschiedenen Molekülmechanik-Modellen bis hin zu semiempirischen, quantenmechanischen und dichtefunktionalen Methoden reichen. Des Weiteren sind im Gegensatz zu anderen integrierten Paketen wie SYBYL die binären Datenformate dokumentiert. So dient das snp-Format in dieser Arbeit als Datenformat für die NEURO-Programmsuite, um Isomerenensembles zu speichern.

Grundsätzlich überwiegen im Rahmen dieser Arbeit die Vorteile der extensiven Dokumentation und der flexiblen Schnittstelle für eine Vielzahl unterschiedlicher Methoden die Nachteile.

Insbesondere die (mangelhafte) Geschwindigkeit und Skalierbarkeit der angebotenen Molekülmechanikmodule machen jedoch den Einsatz weiterer Programme, zumindest zum Generieren der Isomerenensembles, sinnvoll. Die derart erzeugten Daten werden zunächst ins snp-Format überführt und dann auf dem in Kapitel 2 beschriebenen Weg weiterbearbeitet. Das Programmpaket SYBYL<sup>[39]</sup> wird auf diese Weise benutzt.

### 5.1.2 Compiler und Interpreter

Als C++ Compiler kommt für die Programme der NEURO-SUITE die Programmierumgebung *Microsoft Visual C++* (über mehrere Versionen) zum Einsatz. Sämtliche Programme wurden auf die aktuelle Version *Microsoft Visual Studio 2005 – Professional Edition*<sup>[141]</sup> portiert. Die Programme zur Untersuchung der Aromatizität von Annulenen verwenden den C++ Compiler der GNU COMPILER COLLECTION 4.1 (GCC)<sup>[142]</sup> und die Entwicklungsprogramme der CYGWIN<sup>[143]</sup> Programmierumgebung.

Als Skriptsprache kommt im Rahmen dieser Arbeit Tcl/Tk zum Einsatz. Die direkte Steuerung des Programms HYPERCHEM über eine externe Bibliothek (thapi.dll<sup>[144]</sup>) ist damit möglich. Allerdings wird hierzu eine ältere Version von Tcl/Tk (8.0.x) benötigt<sup>[145]</sup>, da es in nachfolgenden Versionen zu einer Veränderung der Aufrufskonvention für externe Bibliotheken gekommen ist (*ABI-breaking change*).

### 5.1.3 Externe Code-Bibliotheken (libraries)

Als externe Bibliotheken werden verwendet:

- Das *HYPERCHEM Application Programming Interface* (HAPI) des Programms HYPERCHEM.<sup>[37],[43]</sup>
- boost::uBLAS<sup>[146]</sup> - Eine Matrix-Bibliothek für die Programmiersprache C++.

### 5.1.4 eigene Code Bibliotheken

#### 5.1.4.1 HYPERSNAP

Die HYPERSNAP-Bibliothek wurde im Rahmen dieser Arbeit entwickelt und ist eine C++ Bibliothek zum bequemen und effizienten Lesen und Schreiben von HYPERCHEM snp-Dateien<sup>[147]</sup>. Dies hat den Vorteil, vom Programm HYPERCHEM produzierte MM- und MC-Rechnungen direkt als Isomerendatenbank lesen zu können. Auch für die NEURO-SUITE eignen sich snp-Dateien als wohldefiniertes binäres Format, da die Zugriffszeiten um Größenordnungen niedriger liegen als bei Textformaten wie z.B. dem xyz-Format.

Das zentrale Datenelement ist die Klasse `data`, welche alle Daten einer snp-Datei zusammenfasst.

```
//                               Module Data
// Module Data has as elements the data objects necessary
```

```
// to completely represent the data of a Snapshotfile
//
// class hirarchy:
//
//           "Data.h"                               "Snp.h"
//
//                                     / fs_Header
//                                     |
//                                     | vs_Header
//                                     <----+
//                                     | s_CooVelo
//                                     |
// Data,                               \ s_Trailer
//   ^         ^
// |         |
// Data_Generator   DDE_Data
```

Die abgeleitete Klasse `DDE_Data` ist erweitert um die Fähigkeit, sich mit einer laufenden HYPERCHEM Instanz zu verbinden und Daten auszutauschen bzw. Rechnungen zu starten.

Die Klasse `Data_Generator` implementiert die Fähigkeit, neue `snp`-Dateien zu generieren, inklusive Konnektivitäten und *variable sized headers*. Das wichtigste Interface der Bibliothek besteht in den *file reader* und *file writer* Klassen, schematisch dargestellt im folgenden Ausschnitt der *header* Datei.

```
// Module iofsnap encapsulates the ability to read and write
// data from/to files.
// class hirarchy:
//
//           Data,   <--- links with pointer to---   fsnapbase
//           ^
//           |
//           |   ofsnap   ifsnap
//           |   \       /
//           |   fsnap
// { DDE_Data }
```

Die HYPERSNAP-Bibliothek ist im Quellcodeordner „`./src/cpp/Hyperchem`“ des *supplementary materials* auf DVD zu finden.

#### 5.1.4.2 Net

Die Klasse `Net` aggregiert sämtliche für ein Kohonennetz der NEURO-SUITE benötigten Daten und implementiert diverse Algorithmen, welche von den einzelnen Programmen der NEURO-SUITE verwendet werden.

Die `Net`-Bibliothek ist im Quellcodeordner „`./src/cpp/Net`“ des *supplementary materials* auf DVD zu finden.

## 5.2 Hilfsskripte für HYPERCHEM

Es folgt eine Kurzbeschreibung der verwendeten eigenen Tcl/Tk-Skripte. Komplexere Skripte sind in der Regel ausführlich in englischer Sprache dokumentiert. Diese Dokumentation wird hier wörtlich übernommen, da eine



Übersetzung lediglich die Gefahr zusätzlicher Übersetzungsfehler bietet und keinen erkennbaren Nutzen bringt. Der Aufruf erfolgt in der Regel durch Aufruf des Menüpunktes (*Script/All Scripts ...* bzw. `SCRIPTS.TK`). Alle Skripte werden als Untermenge der `KOEHLI_SCRIPTS` mit einem *Setup*-Programm installiert.

### 5.2.1 *FIXROT.TK*

fixrot.tk  
(C) 2006 Felix Köhler

#### Synopsis:

Fixrot.tk is a GUI module for the program HyperChem, designed to fix a big flaw of HyperChems simulation engine. Small numerical errors or perhaps imperfections of the random number generator can sum up during heat bath simulations and result in large quantities of the systems energy bound in the rotation of the complete molecular system (molecular dynamics) or translation (Langevin dynamics).

This is highly undesirable. The rotation will distort the structure of molecules due to centrifugal forces. Rotation and translation will account for much of the kinetic energy and result in the real temperature of the system being significantly lower than nominal.

Fixrot.tk solves these problems by acting as a driver for molecular simulations. It breaks up a long simulation (MD, langevin dyn., Monte Carlo) into several small ones having the name `<snapfile>1.snp`, `<snapfile>2.snp`, ... `<snapfile>n.snp`. Between simulations it annihilates the accumulated total rotation energy as well as translation energy of the system and starts the next small simulation, where heat bath coupling will reinject the "lost energy" in randomized form.

It does so by writing an input file and calling "fix\_rotation". The needed corrections to the velocity vectors are read from the the output file and applied to the molecular system within HyperChem.

The atomic mass is an important constant needed for the calculations. There are several peculiarities with HyperChems way to treat atomic masses, which make it necessary to manually assign each mass.

- HyperChem 6.x only uses the correct mass in MD/MM calculations, if an atom has been assigned a forcefield atom type - otherwise mass=1 g/mol.
- even then it uses not the middled molecular weight but the exact isotopic weihts WHILE STILL REPORTING THE MOLECULAR WEIGHTS in the snapshot file.
- The calculated values lag behind the values calculated from HyperChem, notably  $E_{kin}$ . This is due to HyperChem using an integrator of higher order, probably the Verlett leapfrog method. This shouldn't be of importance, as both

the angular momentum and the momentum should not vary (much) during a MD-simulation.

The small simulations are finally joined to one big one by using the "jsnaps" command line program.

Usage:

Load fixrot.tk into the HyperChem Script menu with "load script ...". Use it to start all long simulations.

### 5.2.2 Grid.tcl

*Dihedral driver script für energy surface scans.* Weitere Freiheitsgrade können einfach eingearbeitet werden. Das Ablaufschema sieht dabei aus wie folgt:

- Ermittelt rekursiv die Sollwerte der Gridvariablen.
- Setzt die Sollwerte dieser Variablen als *restraints*.
- Steuert das Programm HYPERCHEM bei einer *restrained* Minimierung.
- Ermittelt die *unrestrained* Energie der erhaltenen Struktur.
- Speichert die Geometrie samt Energie in einer snp-Datei unter Verwendung von POKESNAP.EXE (5.3.15).

### 5.2.3 HCL.TK

HCL.TK ist ein Skript mit grafischer Oberfläche zum Bearbeiten und Testen von Befehlen und Skripten in der *HYPERCHEM Command Language* (HCL). Es wird mitgeliefert mit HYPERCHEM und ist in modifizierter Form ein Teil des angepassten Skriptmenüs in HYPERCHEM.

### 5.2.4 LINTORS.TK (LINEAR TORSIONS)

Aus einer Selektion von N Atomen in HYPERCHEM erzeugt dieses Skript eine Sequenz der Diederwinkel zwischen den angegebenen Atomen in der Reihenfolge der Auswahl der Atome. Diese Diederwinkel werden in einer HYPERCHEM hin-Datei als *selections* { t<sub>1</sub>, t<sub>2</sub>, t<sub>3</sub>, ..., t<sub>N-3</sub> } definiert und können damit von HYPERCHEM verwendet und gespeichert werden.

### 5.2.5 MAKE\_DISTANCE\_MATRIX.TK

Aus einer Selektion von N Atomen in HYPERCHEM erzeugt dieses Skript eine Distanzmatrix zwischen allen selektierten Atomen. Diese Distanzen werden in HYPERCHEM als *selections* { d<sub>1</sub>, d<sub>2</sub>, d<sub>3</sub>, ... d<sub>(N-1)}</sub> definiert und können damit von HYPERCHEM verwendet und gespeichert werden.

### 5.2.6 PRINT\_SELECTION\_VALUES\_GUI.TK

Dieses grafische Skript fragt die Diederwinkel *selections* in einer HYPERCHEM Datei ab, gibt sie aus und führt eine Einteilung in *gauche*<sup>+</sup>, *gauche*, und *anti* analog zu CLASSIFY\_SNP\_CONFORMATIONS.EXE durch. Es wird über das HYPERCHEM Skript-Menü aufgerufen.

### 5.2.7 RELOAD.TK

RELOAD.TK lädt die gegenwärtige Datei in HYPERCHEM neu. In Verbindung mit häufigem Speichern ist dies nützlich, da HYPERCHEM keine Undo-Funktion besitzt. Das Skript stellt einen Teil des modifizieren Skriptmenüs in HYPERCHEM dar.

### 5.2.8 RINGTORS.TK (RING TORSIONS)

Aus einer Selektion von N Atomen in HYPERCHEM erzeugt dieses Skript eine Sequenz der Diederwinkel zwischen den angegebenen Atomen in der Reihenfolge der Auswahl der Atome. Beginn und Ende der Auswahl werden als verknüpft betrachtet. Diese Diederwinkel werden in HYPERCHEM als *selections* {t<sub>1</sub>, t<sub>2</sub>, t<sub>3</sub>, ..., t<sub>N</sub>} definiert und können damit von HYPERCHEM verwendet und gespeichert werden.

### 5.2.9 SCRIPTS.TK

SCRIPTS.TK erzeugt ein Menü mit einem Eintrag für sämtliche Tcl/Tk Skripte mit der Endung .tk im Verzeichnis der KOEHLI\_SCRIPTS.

### 5.2.10 SCRIPT\_LOADER.TK

SCRIPT\_LOADER.TK lädt menügesteuert ein Tcl/Tk Skript mit Anbindung an HYPERCHEM (thapi.dll) in die HYPERCHEM Menüstruktur.

### 5.2.11 SYSTEMATIC\_SEARCH\_PERBROMO\_HBC.TK

Skript zur systematischen Generierung der Hyperfläche des perbromo-HBCs (siehe Kapitel 2.8.9.1). Variiert die *bay*-Diederwinkel, wobei im Rahmen der Veränderung der Diederwinkel benachbarte Bindungen gebrochen und geschlossen werden.

### 5.2.12 TCL.TK

TCL.TK ist ein Skript mit grafischer Oberfläche zum Bearbeiten und Testen der Tcl/Tk Anbindung des Programms HYPERCHEM. Mitgeliefert mit HYPERCHEM und Teil des modifizierten Skriptmenüs in HYPERCHEM.

### 5.2.13 XYZ2COM.TCL

xyz2com.tcl: (C) Felix Koehler 2004

reads coordinates from an xyz - (Multistructure) - File and produces Gaussian input files, that will can be used to optimize each of these structures in a batch.

Usage: tclsh xyz2com.tcl {options} xyzfile

Options (all values have defaults):

```
-k `keywordline`      : keywordline for gaussian input files
-c `charge`           : charge for gaussian input files
-m `multiplicity`     : multiplicity for gaussian input files
-g `gaussiancall`    : command to start the gaussian program
-f1 `singlefixedfile` : read a selectionfile containing all internal
                        coordinates to remain fixed during optimization
                        requires modredundant in keyword line
-f2 `multifixedfile` :
                        read a file with coordinates to remain fixed during
                        optimization. One entry enclosed in
                        STARTFIXED
                        #
                        ...
                        ...
                        ENDFIXED
                        per structure of xyzfile
-s                     : seq_from_xyz sequence nr from commentline of xyzfile
```

Example:

```
tclsh xyz2com.tcl -k "pm3 sp" -c 1 -m 1 -g g03 -f 1-10 a.xyz
resulting comfiles will be called a.1.com a.2.com ...
```

## 5.3 Eigene Programme

Es folgt eine Kurzbeschreibung der verwendeten eigenen Programme. Jedes Kommandozeilenprogramm gibt einen Hilfetext in englischer Sprache aus, der hier wörtlich zitiert wird. Alle Programme werden als Untermenge der KOEHLI\_PROGS mit einem *Setup*-Programm installiert.

### 5.3.1 ANALYSE.EXE [NEURO-SUITE]

Das Programm implementiert das PARAMETRIX-Modul. Es extrahiert die zu trainierenden internen Molekül-Koordinaten aus einer HYPERCHEM hin-Datei und ermittelt den Wertebereich dieser Trainingsparameter für die Normalisierung.

ANALYSE 1.31 2006 F.Koehler: added Angles, Torsions

usage: analyse {options} <snapfilename>

Synapsis: Scans a complete snapshotfile and produces a .dis file containing an entry for each internal parameter, containing the parameter range. Selections are explicitly defined within a .hin file, otherwise the full distance matrix will be scanned.

Options:

```
-v-   verbose off (standard=on)
-l-   short output (standard=long)
-e-   turn off named selections from hinfile
      (standard=on)
```

```

-m$###.## minimum value (real)
  d : distance
  a : angle
  t : torsion
-M$###.## Maximum value (real)
  d : distance
  a : angle
  t : torsion
-d<distancefile>
-h<hinfile> containing selections
-s### number of slots (default=0)
-? display this help

```

attention: named selections beginning with an r or R are considered to be restraints only and are not evaluated.

### 5.3.2 ASSIGN.EXE [NEURO-SUITE]

Assign.exe implementiert die COLORIX Module der NEURO-SUITE. Es extrahiert die Energien von Prototypstrukturen in eine eng-Datei zwecks Darstellung der Energielandschaft durch das Programm NEURO.

Assign 2.1 by Felix Koehler 2006

Usage: assign [options]

options:

```

{ -d<#delta> } float value: |dist1-dist2|<=delta -> 0
{ -z<#maxrelaxation> } int value
{ -f<flag>+/-          default
  a : potential energy      on
  s : stretch energy      off
  b : bending energy       off
  t : torsional energy     off
  n : nonbond energy       off
  v : gradient X (only trained dist.) on
  w : gradient Y (only trained dist.) on
  z : gradient X (all dist.) off
  y : gradient Y (all dist.) off
  o : optimized            on
  p : unoptimized (from snapshotfile) on
}
-n<netfile> (*.net)
-l<logfile> (*.slg)
-s<scrfile> (*.scr) (needs version 1.1)
-t<disfile> (*.dis) trained distances
-p<snpfile> (*.snp) not optimized snapshotfile
-e<engfile> (*.eng)

```

### 5.3.3 CLASSIFY\_BUTAN\_SNP.EXE

Zur Klassifizierung der Butankonformationen in einer snp-Datei als *gauche*<sup>+</sup>, *gauche*, *anti* dient das Programm CLASSIFY\_BUTAN\_SNP. Dies ist eine einfache Applikation, welche geeignet ist, die Verwendung der HYPERSNAP-Bibliothek zu demonstrieren.

classify\_butan\_snp (c) 2006 Felix Koehler

Synopsis: counts the conformations of the butan molecule in a given snpfile.

Usage: `classify_butan_snp <snpfile>`

### 5.3.4 CLASSIFY\_PENTAN\_SNP.EXE

Zur Klassifizierung der Pentankonformationen in einer snp-Datei als  $g^+g^+$ ,  $g^+g^-$ ,  $g^+a$ ,  $g^-g^+$ ,  $g^-g^-$ ,  $g^-a$ ,  $ag^+$ ,  $ag^-$ ,  $aa$  dient das Programm CLASSIFY\_PENTAN\_SNP.

```

classify_pentan_snp (c) 2006 Felix Koehler
Synopsis: counts the conformations of the pentane molecule
          in a given snpfile.
Usage:   classify_pentan_snp <snpfile>

```

### 5.3.5 CLASSIFY\_SNP\_CONFORMATIONS.EXE

Als Weiterentwicklung der beiden obigen Programme beherrscht `classify_snp_conformations.exe` die konformelle Klassifizierung beliebiger Moleküle, wobei die angegebenen Diederwinkel zur Klassifizierung herangezogen werden.

```

classify_snp_conformations (c) 2007 Felix Koehler
Synopsis: counts the conformations in a given snpfile. Does so by
          classifying a list of given dieder angles by configuration.
          tors is of range 0..2Pi
          (a) anti      : 180°+-60° -> Pi +- Pi/3
          (G) gauche+   :  60°+-60° -> Pi/3 +- Pi/3
          (g) gauche-   : -60°+-60° -> 5/3Pi +- Pi/3
Usage:   classify_snp_conformations <selfile> <snpfile>
          <selfile> : name of selection file with dieder definitions
          <snpfile> : name of snapshotfile to analyze.

```

### 5.3.6 CYCHEX\_TORS.EXE

CYCHEX\_TORS.EXE führt eine Gridrechnung über den Konformationsraum des Cyclohexans durch. Dabei werden zwei gegenüberliegende Diederwinkel in den Grenzen  $[-80^\circ \dots 80^\circ]$  variiert, anschließend wird ein dritter Diederwinkel variiert, während gleichzeitig die beiden vorherigen Diederwinkel auf  $0^\circ$  festgehalten werden.

### 5.3.7 EXAKTE\_LOESUNG.EXE

Das Programm berechnet die Anzahl der theoretisch denkbaren CANE-Strings eines [n]Annulens.

```

Usage:
exakte_loesung.exe Anzahl_der_Bindungen

```

### 5.3.8 FIX\_ROTATION.EXE

```

fix_rotation 1.0
(C) 2006 by Felix Koehler
Synopsis:
Takes the masses and velocities from an inputfile in format
mass[i] vx vy vz

```

```

... .. ..
and writes the correctional velocities, which have to be
added to the velocities to get a system without any net translation
and without any net rotation into the outputfile

```

Format of outputfile:

```

  x    y    z    vx    vy    vz
  ..   ..   ..   ..   ..   ..

```

Options:

```

t- : no translational correction
r- : no rotational correction

```

Usage:

```

rot_control -i<inputfile> -o<outputfile> {-options}

```

### 5.3.9 GENERATE\_SNAP\_HEADERS

Erzeugt aus einer hin-Datei eine snp-Datei, welche als Schablone dient und mit Isomerendaten befüllt werden kann.

```

generate_snap_headers

```

```

(c) 2003 Felix Koehler

```

Usage: generate\_snap\_headers <...>

Options:

```

-s<snapname> name of the hyperchem snapshot file to produce.
-i<hinname>  name of the hyperchem file to construct the snapshot header for

```

### 5.3.10 HCS2SNP.EXE

Das Programm konvertiert die Strukturen aus einer HYPERCHEM Konformati-  
onssuche<sup>[37],[43]</sup> in das snp-Format,<sup>[147]</sup> um anschließende Rechnungen damit  
durchführen zu können.

Usage: hcs2snp hcsfile snpfile

```

Converts a HYPERCHEM Conformational search file into a
Snapshotfile. The Snapshotfile must preexist and feature
a valid header.

```

### 5.3.11 JSNAPS.EXE

```

jsnaps 0.3: (C) 2006 Felix Koehler

```

Synopsis:

```

joins single snapshotfiles given as <stub>1, <stub>2, <stub>3 to
one large snapshotfile named stub.snp. Incomplete series can be
handled by explicitly specifying the range to join with the -a and
-z options

```

Usage: jsnaps {options} <stub>

```

-a <#first> : specify first snapshot number
-z <#last>  : specify last snapshot number

```

Example 1: 'jsnaps dynamik'

```

joins the files dynamik1.snp, dynamik2.snp, ..., dynamikxxx.snp
where xxx is last number in a continuous series of snp-files.

```

Example 2: 'jsnaps -a 0 -z 200 dynamik'

```

joins all snp-files dynamikxxx.snp with xxx=0..200 even if some
numbers are missing.

```

### 5.3.12 MAKE\_2ND [NEURO-SUITE]

MAKE\_2ND bereitet die MINIX2 und COLORIX2 Module vor.

make\_2nd 1.0 (C) 2006 Felix Koehler  
Part of the Neuro-Suite.

-----  
Usage: make\_2nd <stub1> <stub2>  
    Produces from the given <stub1>.net the files  
    <stub2>.net, <stub2>.scr necessary for a secondary minimization  
    protocol (MINIX2).

### 5.3.13 MINIMIZE.EXE [NEURO-SUITE]

MINIMIZE.EXE implementiert die MINIX-Module der NEURO-Suite. Dabei verwendet es das Programm HYPERCHEM zum Durchführen der eigentlichen Rechnungen.

Minimize 3.4 : (C) 2006 Felix Koehler  
Usage : minimize <options...> <generic filename>  
    -D+(debug mode, default -D-)  
    -d<disfile>  
    -s<snapfile>  
    -l<logfile>  
    -S# (Use source # for starting geometry)  
        1: (default) source snapshot  
        2: target as source (reminimize with tighter convergence)  
    -M- minimize off (default = on just does single points for quick prescan)  
    -m<hinfile>  
    -n<netfile>  
    -c<scriptfile>  
    -o<optimization maximum cycles>  
    -z<maxRelaxtion (=0)>  
    -x<convergence\_limit>  
    -g starting\_geometry  
        n: from net (net-parameters as *restraints*)  
        s: from snapshot (otherwise default)  
    -R- Repair mode off (default = on)  
        if a \*.slg.log file exists, minimizes only structures where  
        the field converged is 'n'. Incomplete Minimize Runs are completed.  
    -r (use restraint)  
        d# (distances from \*.dis with force #)  
        a# (angles from \*.dis with force #)  
        t# (torsions from \*.dis with force #)  
        o-/+ (others from \*.hin, default = true)  
    -p<pathname of HYPERCHEM5> - if not given trys to load  
        running process

#### Description:

minimizes the structures given in the scriptfile / netfile  
and saves the produced structures in the logfile \*.slg  
Scriptnames normally result from the Netfilenames.

#### New:

- Scripts are not overwritten, the success of the optimization is logged in the file \*.slg.log.
- Old Logfilenames \*.log are automatically converted to \*.slg.
- starting geometry pure snapshotfile

### 5.3.14 NEURO.EXE [NEURO-SUITE]

NEURO.EXE dient als *Frontend*-Programm für die Kommandozeilenprogramme der NEURO-Suite. Die Bedienung erfolgt über eine grafische Oberfläche und



wird in den Tutorials in Kapitel 2.8.1 beschrieben. NEURO.EXE erzeugt dabei die teils recht komplexen Kommandozeilen zum Aufruf der übrigen Programme der Suite. Eine Logdatei der getätigten Aktionen wird im Verzeichnis %TEMP% unter dem Dateinamen Neuro#####.log erstellt, wobei ##### für eine Zufallszahl steht. Diese Datei enthält obige komplexe Kommandozeilen, falls Bedarf nach „manueller“ Nachbesserung auf der Kommandozeile besteht.

### 5.3.15 NN.EXE [NEURO-SUITE]

NN.EXE implementiert die Module KOHONIX und PROTOTYPIX und führt das Training einer Kohonenkarte inklusive Ermittlung der Prototypstrukturen durch.

NN 3.0 (c) 2006 Felix Koehler

Usage: nn <...>

Options:

```
{ -?          display this help screen }
{ -A#         maximum learning value (default 0.15)}
{ -D#         maximum learning distance (default 0)
               0: 0.4 * Max(NETX, NETY)}
{ -N<file>    name of the neighborhood function file
{ -a#         assignmode 0: neurons choose snap (default)
               1: snaps choose neuron}
{ -b+         output the net with the lowest distance_sum during training
               (default = false)}
{ -dDelta     allowable error in percent (when assigning neurons)
               (default=2.0) }
{ -e#         int #: repeat training count (default=5) }
{ -f+         fixed number of epochs (Standard = -)
{ -h<file>    name of HYPERCHEM hin-file, in this version not needed
{ -i<file>    name of distance-file
{ -l<file>    name of binary .sav file to read and scan (obsolete) }
{ -m<file>    name of HYPERCHEM script-file
{ -n<file>    name of log-file to generate (not yet implemented)
{ -o<file>    name of the Ascii net-file to be output
{ -p<file>    name of HYPERCHEM snapshot-file
{ -q-         quadratic learning decay
{ -r#         int #: random init (default = 0) }
{ -s<binary>  name of binary .sav file to generate (default=NET.sav)}
{ -v#         int #: verbose mode (default= 1) }
{ -x#         size of net in x direction (default=25) }
{ -y#         size of net in x direction (default=25) }
{ -z#         int #: minimize structure relaxation (default = 0) }
```

### 5.3.16 POKESNAP.EXE

POKESNAP.EXE implementiert das Schreiben von Geometriedaten an eine beliebige Position in eine snp-Datei und ist nützlich für diverse HCL-basierte Skripte.

```
#####
#          pokesnap 1.0 (c) 2006 Felix Koehler          #
# Reads coordinates of a HYPERCHEM HIN-File and writes them in a      #
# structure given by number in an existing HYPERCHEM Snapshot-File.   #
# Usage: pokesnap <...>                                           #
# Options:                                                         #
# -n#          int #: target number to write structure in snap      #
# -h<hin>      name of HYPERCHEM hin-file                        #
# -s<snap>     name of HYPERCHEM snapshot-file                  #
# -e<erg>      potential energy of the target snap              #
#####
```

### 5.3.17 SEQLOG2XYZ.EXE

seqlog2xyz 1.0.2

Usage: seqlog2xyz STUB {-option, ...}

Converts multiple logfiles starting with STUB1, STUB2, ...  
into an xyz file.

The sequence number is written into the comment.

Gaps in sequence are possible.

Copyright Felix Koehler 2004,2007

Options:

-h/-?: this help

-s: suffix

-a: Start index (default = 1)

-j: filestub for logsequence with indene-derivates

-k: filestub for logsequence with isoinden-derivates

-m: filestub for the logsequence containing the magnetical  
susceptibility. log files should be obtained by CSGT method.

-n: filestub for the logsequence containing the NICS values.  
logfiles should be obtained by GIAO method.

-z: Stop index (default=10000)

-V: print version info

Hints: Providing the dot in STUB and suffix is usually necessary, eg.:

seqlog2xyz conformer\_database. -a 1 -z 10000 -s .out

### 5.3.18 SNP2XYZ.EXE

Usage: snp2xyz <snpfile> <xyzfile> {-option, ...}

Converts a HYPERCHEM Snapshotfile into a multiple structures  
xyz file.

Options:

-h/-?: this help

-p: Potential Energy in comments on(default off)

-t: Total Energy in comments on(default off)

-a: Start index

-z: Stop index

-s: separate structures into single files: (default = off)  
xyzfile0.xyz, xyzfile1.xyz, xyzfile2.xyz, ...

### 5.3.19 SUBSNAP.EXE

subsnap 0.3 : (C) 2006 Felix Koehler

Usage : subsnap <options...>

Extracts only part of the molecular system and/or only part of the  
time series from a snapshot file. Produces a new snapshotfile  
containing this subsnap.

The atoms are given as a whitespace separated list.

Make sure the list argument stays ONE argument by quoting "..."

Options:

-s<snapfile> : Name of source snapshotfile

```

-t<snapfile>      : Name of target snapshotfile
-a<#first>       : specify first snapshot number (default 0)
-z<#last>        : specify last snapshot number (default -1=all)
-i<#intervall>   : only extract every intervall'th timestep.
-?              : display this help
"-l#1 #2 #3 ..." : White space separated list of pivot atoms

```

### 5.3.20 TRANSLATE\_SNAP.EXE

Usage: translate\_snap <snpfile> {-option, ...}  
 Translates all atoms in each step of a dynamics, so that the centre of mass or, if given, a central atom is set translated to the origin.  
 If a periodic box is used, the atoms are rearranged into the box

Options:

```

-h/-?: this help
-x#: x-coordinate of the origin
-y#: y-coordinate of the origin
-z#: z-coordinate of the origin
{-n#: Nr of central atom in snapshotfile}

```

### 5.3.21 XYZ2CANE.EXE

xyz2cane 1.0.1

(c) 2006, 2007 Felix Koehler

Synopsis:

Writes a (CANE) Canonical Enumeration nomenclature string list for all structures in xyzfile.

Usage: xyz2cane <Options>

Options:

```

-s <selectionname> name of the selection file
-d : don't delete double cane values with same energy
-i <xyzname> name of the xyz input file
-n <nicsname> name of the xyz file with nics positions as ghost atoms
    (Bq)
-o <cleanxyzname> name of the xyz outputfile, doublettes cleaned out
-m <mutatealgorithm>: Produce xyzfiles with inden and isoinden derivatives
    into structure. 8 is standard algorithm.
    1: substitute preferably at outer atoms
    ....
    8: substitute preferably at most planar section
    1: ~ centerdistance
    2: ~ planarity * centerdistance
    3: ~ planarity * scaled centerdistance
    4: ~ planarity^2 * centerdistance
    5: ~ planarity^2 * scaled centerdistance
    6: ~ planarity^3 * centerdistance
    7: ~ planarity^3 * scaled centerdistance
    8: ~ planarity
-x <maxbondlength> : only in combination with -m option. Maximum
    allowed bondlength for substitution. Default is
    1.40, which is larger than the aromatic bond in benzene
    Use this to make sure isoinden/indene substitutioun is
    never at single bonds.
-v : verbose output
-V : output version and exit
-vv : very verbose output, prints distances and dieters
-c : comments contain alternate number for structure
-e <energy> : energy difference where structures still are considered
    to be of same energy, default = 0.05 kcal/mol.

```

### 5.3.22 XYZ2SNP.EXE

#### Synopsis:

Converts an xyz multistructure file into a HYPERCHEM snapshotfile.

A template HYPERCHEM hin file corresponding to the exact structure of the first molecule entry of the xyz file must be given. It is needed for information about connectivity and atom types.

The cartesian positions of the atoms in first xyz entry and the HYPERCHEM template file must be identical to allow for correct renumbering from xyzfile to snapshotfile.

Usage: xyz2snp <Options>

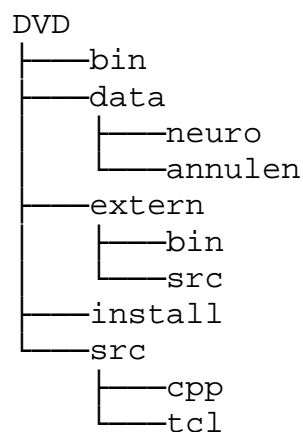
#### Options:

- i <xyzname> name of the xyz input file
- o <snpname> name of the snp output file
- t <hinname> name of the HYPERCHEM templatefile
- r use existing snapshotfile headers. Behavior is undefined if snapshotfile is incompatible. (default=off)

## 6 Appendix B: Distribution und supplementary material

Eine Distribution der erstellten Programme auf DVD ist dieser Arbeit beigelegt. Um eine einfache Reproduktion der Arbeitsumgebung zu erhalten, sind auch einige Programme und Hilfsmittel aus Händen Dritter ausführbar und im Quelltext beigelegt.

Die Verzeichnisstruktur auf der DVD wird im Folgenden kurz erläutert. Eine detaillierte Installationsanweisung ist in der Datei INSTALL.TXT im Stammverzeichnis der DVD zu finden.



**Abbildung 206:** Verzeichnisstruktur der beigelegten DVD

- /bin Enthält die ausführbaren Dateien für MICROSOFT WINDOWS XP. Die Programme können nach Anwendung der Installationsanleitung in INSTALL.TXT verwendet werden.
- /extern Programme aus Händen Dritter.
- /extern/src Enthält Unterverzeichnisse mit Quelltext von Programmen aus Händen Dritter im Ursprungszustand, wie vom Originalautor ausgeliefert. Sollten diese mit Änderungen verwendet werden, ist eine *patch*-Datei mit den Abweichungen mitgeliefert, welche den Zustand der hier verwendeten Version mit Hilfe der GNU DIFFUTILS wiederherstellt.<sup>[148]</sup> In dem Fall beschreibt eine README Datei im übergeordneten Ordner die vorzunehmenden Schritte.
- /extern/bin Enthält ausführbare Dateien aus Händen Dritter. Können behandelt werden wie Dateien im Ordner /bin.
- /install Enthält Unterordner mit automatischem Installer (eigene Programme und Fremdquellen).
- /data/neuro *supplementary material* zu Kapitel 2

/data/annulen *supplementary material* zu Kapitel 3

/src Quellcode selbst erstellter Programme.

/src/cpp Enthält Unterverzeichnisse für Programme in der Programmiersprache C++. Wenn zum Erstellen eines Programms der Microsoft-Compiler<sup>[141]</sup> verwendet wird, enthält das Verzeichnis eine Projektdatei mit der Endung `.vcproj`. Wird zum Erstellen der GNU-Compiler<sup>[142]</sup> `g++` verwendet, enthält das Verzeichnis eine Datei mit Namen `Makefile`. Erstere Variante ergibt in der Regel ausführbare Dateien plattformabhängig für Microsoft Windows, letztere Variante für GNU/Linux und Windows/Cygwin.<sup>[143]</sup>

/src/tcl Enthält Programme in der plattformunabhängigen Skriptsprache Tcl/Tk.<sup>[145]</sup>

## 7 Appendix C: Anmerkungen und Literaturverzeichnis

- [1] C. S. Wannere, P. v. R. Schleyer, *Org. Lett.* **2003**, *5*, 865-868.
- [2] GAUSSIAN 03 *Revision C.02*, M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery, Jr., T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez, J. A. Pople, Gaussian, Inc., Wallingford CT **2004**.
- [3] C. Priesner, K. Figala, *Alchemie* 1. Aufl., Verlag C.H. Beck, München **1998**.
- [4] J. Dalton, *A New System of Chemical Philosophy* Manchester **1808**.
- [5] J. D. Watson, *Die Doppel – Helix* 19. überarb. Auflage, Rowohlt Taschenbuch Verlag, Regensburg **1993**.
- [6] M. Ela M, F. H. White jr., C. B. Anfinsen, *Science* **1957**, *125*, 691-692.
- [7] C. B. Anfinsen, *Science* **1973**, *181*, 223-230.
- [8] C. B. Anfinsen, H. A. Scheraga, *Adv. Protein Chem.* **1975**, *29*, 205-300.
- [9] T. Wainwright, B. J. Alder, *Nuovo Cimento* **1958**, *9 (Suppl. No. 1)*, 116-132.
- [10] B. J. Alder, T. E. Wainwright, *J. Chem. Phys.* **1959**, *31*, 459-466.
- [11] B. J. Alder, T. E. Wainwright, *J Chem. Phys.* **1960**, *33*, 1439-1451.
- [12] Abbildung entnommen aus E. Walker, *Lecture Notes* **2005**.
- [13] D. Stock, A.G.W. Leslie, J.E. Walker, *Science* **1999**, *286*, 1700-1705.
- [14] J. E. Walker, *Angew. Chem.* **1998**, *110*, 2439-2450.
- [15] RCSB – eine frei zugängliche Internetdatenbank für Proteinstrukturen samt der zur Lösung der Struktur verwendeten spektroskopischen Daten, (Röntgenstruktur, NMR, etc), <http://www.rcsb.org>.
- [16] CCDC - frei zugängliche Internetdatenbank für Molekül- und Kristallstrukturen gewonnen per Röntgenstrukturanalyse und Neutronendiffraktion an Einkristallen und Pulvern, <http://www.ccdc.cam.ac.uk>.
- [17] R. C. v. Schaik, W. F. v. Gunsteren, H. J. C. Berendsen, *J. Comp.-Aid. Mol. Design* **1992**, *6*, 97-112.

- [18] M. R. Garey, D. S. Johnson, *Computers and Intractability: A guide to the Theory of NP-Completeness*, Freeman, San Francisco **1979**.
- [19] K.-H. Zimmermann, *Comp. Phys Comm.* **2002**, *144*, 297-309.
- [20] D. S. Abrams, S. Lloyd, *Phys. Rev. Lett.* **1998**, *81*, 3992-3995.
- [21] J. C. Christopher, *Essentials of Computational Chemistry – Theories and Models*, Second Ed., Wiley, Chichester, **2004**, 69-71
- [22] A. E. Howard, P. A. Kollman, *J. Med. Chem.* **1988**, *31*, 1669-1675.
- [23] T. C. Beutler, W. F. van Gunsteren, *Chem. Phys. Lett.* **1995**, *237*, 308-316.
- [24] V. Krautler, W. F. van Gunsteren, P. H. Hunenberger, *J. Comput. Chem.* **2001**, *22*, 501-508.
- [25] A. E. Howard, P. A. Kollman, *J. Med. Chem.* **1988**, *31*, 1669-1675.
- [26] E. L. Eliel, S. H. Wilen, L. N. Mander, *Stereochemistry of Organic Compounds*, Wiley/VCH, Weinheim **1994**, p. 600.
- [27] M. Saunders, K. N. Houk, Yun Dong Wu, W. C. Still, M. Lipton, G. Chang, W. C. Guida, *JACS* **1990**, *112*, 1419-1427.
- [28] W. F. v. Gunsteren, H. J. C. Berendsen, *Angew. Chem.* **1990**, *102*, 1020-1055.
- [29] S. J. Weiner, P. A. Kollman, D. T. Nguyen, D. A. Case, *J. Comput. Chem.* **1986**, *7*, 230-252.
- [30] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus, *J. Comput. Chem.* **1983**, *4*, 187-217.
- [31] W. F. van Gunsteren, S. R. Billeter, A. A. Eising, P. H. Hunenberger, P. Kruger, A. E. Mark, W. R. P. Scott, I. G. Tironi, *Biomolecular Simulation: The GROMOS96 Manual and User Guide*, VdF: Hochschulverlag AG an der ETH Zurich and BIOMOS b.v, Zurich, Groningen, **1996**.
- [32] N. L. Allinger, *JACS* **1977**, *99*, 8127-8134.
- [33] N. L. Allinger, Y. H. Yuh, J. H. Lii, *JACS* **1989**, *111*, 8551-8566.
- [34] W. L. Jorgensen, J. Tirado-Rives, *JACS* **1988**, *110*, 1657-1666.
- [35] T. A. Halgren, *J. Comput. Chem* **1996**, *17*, 490, 520, 553, 616.
- [36] F. Mohamadi, N. G. Richards, W. C. Guida, R. Liskamp, M. Lipton, C. Caufield, G. Chang, T. Hendrickson, W. C. Still, *J. Comput. Chem.* **1990**, *11*, 440-467.
- [37] HYPERCHEM(TM) *Professional 6.03*, Hypercube, Inc., Gainesville, Florida **2000**.
- [38] M. Clark, R. D. Cramer III., N. van Opdenbosch, *J. Comput. Chem.* **1989**, *10*, 982-1012.
- [39] SYBYL *Version 6.9*, Tripos Inc., St. Louis **2002**
- [40] H. Kessler, C. Griesinger, J. Lautz, A. Mueller, W. F. van Gunsteren, H. J. C. Berendsen, *JACS* **1988**, *110*, 3393-3396.
- [41] N. Metropolis, A. W. Rosenbluth, W. Arianna, M. N. Rosenbluth, A. H. Teller, H. Augusta, E. Teller, *J Chem. Phys.* **1953**, *21*, 1087-1092.
- [42] S. H. Northrup, J. A. McCammon, *Biopolymers* **1980**, *19*, 1001-1016.



- [43] HYPERCHEM(TM) *Professional 7.52*, Hypercube, Inc., Gainesville, Florida **2003**.
- [44] W. F. van Gunsteren, A. E. Mark, *Eur. J. Biochem.* **1992**, *204*, 947-961.
- [45] A. E. Garcia, *Phys. Rev. Lett.* **1992**, *68*, 2696-2699.
- [46] A. Amadei, A. B. M. Linssen, and H. J. C. Berendsen, *Proteins: Struct., Funct., Genet.* **1993**, *17*, 412-425.
- [47] M. F. DaCruz, M. Zimmer, *Inorg. Chem.* **1996**, *35*, 2872-2877.
- [48] M. E. Karpen, D. J. Tobias, and C. L. Brooks III., *Biochemistry* **1993**, *32*, 412-420.
- [49] GOPENMOL *version 3.00*, CSC - Scientific Computing Ltd., Finland, **2005**.
- [50] C. Manetti, M.-A. Ceruso, A. Giuliani, C. L. Webber jr., J. P. Zbilut, *Phys. Rev. E* **1999**, *59*, 992-998.
- [51] J.-P. Eckmann, S. O. Kamphorst, D. Ruelle, *Europhys. Lett.* **1987**, *4*, 973-977.
- [52] J. A. Hartigan, *J. Am. Stat. Assoc.*, **1967**, *62*, 1140-1158.
- [53] M. Levitt, *J. Mol. Bio.* **1983**, *168*, 621-657.
- [54] T. F. Cox, M. A. A. Cox, *Multidimensional Scaling*, CRC Monographs on Statistics & Applied Probability Second Ed., Chapman & Hall **1988**.
- [55] R. O. Duda, P. E. Hart, D. G. Stork, *Pattern Classification*, Second Ed., Wiley-Interscience, **2001**, S. 542.
- [56] R. O. Duda, P. E. Hart, D. G. Stork, *Pattern Classification*, Second Ed., Wiley-Interscience, **2001**, S. 554.
- [57] G. A. Carpenter, S. Grossberg, *Appl. Opt.* **1987**, *26*, 4919.
- [58] J. Zupan, J. Gasteiger, *Neural Networks for Chemists-An Introduction*, VCH, Weinheim **1993**.
- [59] K. O. Alper, C. K. Bagdassarian, *J. Theo. Comput. Chem.* **2002**, *1*, 45-52.
- [60] K. O. Alper, M. Singla, J. L. Stone, C. K. Bagdassarian, *Protein Science*, **2001**, *10*, 1319-1330.
- [61] M. Faraday, *Phil. Trans.*, London **1825**, 440.
- [62] A. Kekulé, *Bull. Soc. Chim.*, Paris **1865**, *3*, 98.
- [63] E. Hückel, *Z. Physik* **1931**, *70*, 310.
- [64] E. Heilbronner, *Tetrahedron Lett.*, **1964**, *29*, 1923.
- [65] F. A. Carey, R. J. Sundberg, H. J. Schäfer, *Organische Chemie*, VCH Wiley, Weinheim **1995**.
- [66] C. S. Wannere, D. Moran, N. L. Allinger, B. A. Hess, Jr., L. J. Schaad, P. v. R. Schleyer, *Org. Lett.* **2003**, *5*, 2983-2986.
- [67] T. M. Krygowski, M. Cyranski, *Tetrahedron* **1996**, *52*, 1713-1722.
- [68] A. Julg, P. Francois, *Theoret. Chim. Acta* **1967**, *7*, 249-259.
- [69] H. J. Dauben, Jr., J. D. Wilson, J. L. Laity, *JACS* **1968**, *90*, 811-812.
- [70] H. J. Dauben, Jr., J. D. Wilson, J. L. Laity, *JACS* **1969**, *91*, 1991-1998.
- [71] W. H. Flygare, *Chem. Rev.* **1974**, *74*, 653-687.
- [72] L. Pauling, *J. Chem. Phys.*, **1936**, *4*, 673.

- [73] P. v. R. Schleyer, C. Maerker, A. Dransfeld, H. Jiao, N. J. Hommes, *JACS* **1996**, *118*, 6317.
- [74] P. v. R. Schleyer, H. Jiao, N. J. R. v. Hommes, V. G. Malkin, O. L. Malkina, *JACS* **1997**, *119*, 12669-12670.
- [75] H. E. Zimmerman, *JACS* **1966**, *88*, 1564-1567.
- [76] H. E. Zimmerman, *Science* **1966**, *153*, 83-144.
- [77] H. E. Zimmerman, *Accts. Chem. Res.* **1971**, *4*, 272-280.
- [78] a) E. Clar, *Polycyclic Hydrocarbons*, Academic Press, New York **1964**. b) E. Clar, *The Aromatic Sextet*, Wiley, London, **1972**.
- [79] R. Herges, D. Geuenich, *J. Phys. Chem. A* **2001**, *105*, 3214-3220.
- [80] D. Geuenich, K. Hess, F. Koehler, R. Herges, *Chem. Rev.* **2005**, *105*, 3758-3772.
- [81] E. Erlenmeyer, *Ann.* **1866**, 327.
- [82] P. Pascal, *Ann. Chim. P.*, **1910**, *19*, 5.
- [83] W. Kutzelnigg, *J. Chem.* **1980**, *19*, 193.
- [84] D. Ajami, O. Oeckler, A. Simon, R. Herges, *Nature* **2003**, *426*, 819-821.
- [85] H. Gretschel, A. F. Möbius, *Archiv der Mathematik und Physik* **1869**, *49*, 1-9.
- [86] A. D. Becke, *J. Chem. Phys.* **1993**, *98*, 5648-5852.
- [87] C. Lee, W. Yang, R. G. Parr, *Phys. Rev. B* **1998**, *37*, 785-789.
- [88] a) H. J. Dauben, Jr., J. D. Wilson, J. L. Laity, *JACS.* **1968**, *90*, 811 und **1969**, *91*, 1991. b) H. J. Dauben, Jr., J. D. Wilson, J. L. Laity, *Diamagnetic Susceptibility Exaltation as Criterion in Non-Benzenoid Aromatics*, Snyder (Ed.) Vol. 2, Academic Press, New York **1971**.
- [89] A. R. Katritzky, P. Barczynski, G. Musumarra, D. Pisano, M. Szafranll, *JACS* **1989**, *111*, 7-15.
- [90] T. M. Krygowski, M. K. Cyranski, *Chem. Rev.* **2001**, *101*, 1385-1419.
- [91] J. Kruszewski, T. M. Krygowski, *Tetrahedron Lett.* **1972**, 3839.
- [92] T. M. Krygowski, *J. Chem. Inf. Comput. Sci.* **1993**, *33*, 70.
- [93] R. F. W. Bader, *Atoms in Molecules*, Clarendon Press, Oxford, U.K., **1990**.
- [94] T. Kohonen, *Appl. Opt.*, **1987**, *26*, 4919-4930.
- [95] G. Raddatz, *Berechnung der Informationsdimension von Datensätzen zur geeigneten Dimensionierung von Kohonenkarten*, Studienarbeit, Uni Tübingen, **1993**.
- [96] K. Pawelzik, H.-U. Bauer, *Informatik Fachberichte*, **1992**, *306*, 146-152.
- [97] H. Shouno, K. Kurata, *Bio. Cybern.* **2001**, *85*, 241-246.
- [98] Im Gegensatz zum *unsupervised learning* sind beim *supervised learning* eine Auswahl an richtigen Ergebnissen für die Ausgabeschicht bekannt. Bei fast allen anderen Netzwerkarten wie der „*Linear Learning Machin*“, dem Hopfield Netzwerk dem Bidirektionalen Assoziativen Speicher (ABAM) oder den *Counter-Propagation*-Netzen wird versucht die Ausgabe des Netzes möglichst genau an diese Zielvorgabe anzupassen.

- [99] D. O. Hebb, *The Organization of Behavior: A Neuropsychological Theory*, Lawrence Erlbaum Associates, **2002**.
- [100] H. J. Ritter, T. M. Martinetz, K. J. Schulten, *Neuronale Netze*, Addison-Wesley, München, **1991**.
- [101] T. Kohonen, *Bio. Cybern.* **1982**, 43, 59-69.
- [102] M. Außendorf, [http://de.wikipedia.org/wiki/Self-Organizing\\_Maps](http://de.wikipedia.org/wiki/Self-Organizing_Maps), Version 17.11.2006.
- [103] Das MM+ Kraftfeld ist eine Hyperchem-spezifische Version des bekannten MM2 Kraftfelds von Allinger *et al.* Das Funktional ist identisch, zusätzlich wurden für universellere Verwendbarkeit neue Parameter von der Firma Hyperchem hinzugefügt.
- [104] Extrapolation von RMSGend(Hexan)  
 $y = a - b/(1 + c*x)^{1/d}$   
 $\text{Chi}^2/\text{DoF} = 8.0346\text{E-}8$   
 $R^2 = 0.99987$   
 $a = 0.24396 \pm 0.00027$   
 $b = 0.05405 \pm 0.00063$   
 $c = 0.00371 \pm 0.0008$   
 $d = 0.23857 \pm 0.04489$ .
- [105] ORIGIN 7.5G SR6, OriginLab Corporation, Northhampton **2006**, <http://www.OriginLab.com>.
- [106] Extrapolation von RMSDSNend(Hexan):  
 $y = \text{START} + (\text{END} - \text{START}) * x^n / (k^n + x^n)$   
 $\text{Chi}^2/\text{DoF} = 4.6143\text{E-}9$   
 $R^2 = 0.99994$   
 $\text{START} = 0.08544 \pm 0.00048$   
 $\text{END} = 0.06482 \pm 0.00007$   
 $k = 28.74735 \pm 1.32173$   
 $n = 1.43383 \pm 0.05537$
- [107] Extrapolation von RMSGend(Heptan):  
Modell: Logistic  
 $y = A2 + (A1-A2)/(1 + (x/x0)^p)$   
 $\text{Chi}^2/\text{DoF} = 9.0257\text{E-}7$   
 $R^2 = 0.99972$   
 $A1 = 0.27017 \pm 0.00177$   
 $A2 = 0.38683 \pm 0.00082$   
 $x0 = 50.06279 \pm 1.44358$   
 $p = 1.70621 \pm 0.08735$ .
- [108] Extrapolation von RMSDSNend(Heptan):  
Modell: Logistic  
 $y = A2 + (A1-A2)/(1 + (x/x0)^p)$   
 $\text{Chi}^2/\text{DoF} = 9.471\text{E-}9$

$$R^2 = 0.99995$$

$$A1 = 0.15536 \pm 0.00056$$

$$A2 = 0.12351 \pm 0.00009$$

$$x0 = 30.32168 \pm 1.04148$$

$$p = 1.57664 \pm 0.0531.$$

[109] Extrapolation von RMSG<sub>end</sub>(Oktan)

Modell: Logistic

$$\text{Equation: } y = A2 + (A1-A2)/(1 + (x/x0)^p)$$

$$\text{Chi}^2/\text{DoF} = 9.4114\text{E-}6$$

$$R^2 = 0.99888$$

$$A1 = 0.31843 \pm 0.00467$$

$$A2 = 0.51274 \pm 0.00587$$

$$x0 = 124.81603 \pm 7.96776$$

$$p = 1.27995 \pm 0.11324.$$

[110] Extrapolation von RMSDSN<sub>end</sub>(Oktan)

Modell: Logistic

$$\text{Equation: } y = A2 + (A1-A2)/(1 + (x/x0)^p)$$

$$\text{Chi}^2/\text{DoF} = 2.3691\text{E-}7$$

$$R^2 = 0.99958$$

$$A1 = 0.23741 \pm 0.00179$$

$$A2 = 0.18029 \pm 0.00095$$

$$x0 = 58.96047 \pm 3.63303$$

$$p = 1.06163 \pm 0.07995.$$

[111] J. Dale, *Stereochemistry and Conformational Analysis*, Universitetsforlaget, Oslo, Norwegen **1982**, S. 689

[112] J. B. Hendrickson, *JACS*, **1967**, 89, 7047-7061.

[113] D. Cremer, J. A. Pople, *JACS* **1975**, 97, 1354-1358.

[114] POV-Ray *Version 3.6.1*, Vision Raytracer Pty. Ltd. **2004**,  
<http://www.povray.org/>.

[115] Extrapolation von RMSG<sub>end</sub>(Cyclohexan)

Modell: Logistic

$$\text{Gleichung: } y = A2 + (A1-A2)/(1 + (x/x0)^p)$$

$$\text{Chi}^2/\text{DoF} = 1.329\text{E-}7$$

$$R^2 = 0.99823$$

$$A1 = 0.12019 \pm 0.002$$

$$A2 = 0.14579 \pm 0.00224$$

$$x0 = 108.58848 \pm 17.5758$$

$$p = 0.73211 \pm 0.15962.$$

[116] Extrapolation von RMSDSN<sub>end</sub>(Cyclohexan)

Modell: Logistic

$$\text{Gleichung: } y = A2 + (A1-A2)/(1 + (x/x0)^p)$$

$$\text{Chi}^2/\text{DoF} = 3.3372\text{E-}8$$

$$R^2 = 0.99909$$

$$A1 = 0.04172 \pm 0.00205$$

$$A2 = 0.02394 \pm 0.00046$$

$$x_0 = 31.54754 \pm 8.35402$$

$$p = 0.87591 \pm 0.14061.$$

- [117]I. Kolossváry, W. C. Guida, *J. Comput. Chem.*, **1993**, *14*, 691.
- [118]HYPERCHEM *Conformational Search* über den Konformationsraum des Dibromcyclooctens, im supplementary material:  
DVD:/data/neuro/dibromcycloocten/SEARCH.HCS.
- [119]K. A. Dill, *Biochemistry* **1985**, *24*, 1501-1509.
- [120]F. Dötz, *Von zwei- zu dreidimensionalen Graphitausschnitten*, <http://deposit.ddb.de/cgi-bin/dokserv?idn=960727043>, Dissertation: Universität Mainz **2000**.
- [121]HYPERCHEM *Conformational Search* über den Konformationsraum des Permethoxy-HBC, supplementary material:  
DVD:/data/neuro/permethoxy-HBC/CONFSEARCH.HCS
- [122]G. Schröder, *Cyclooctatetraene*, VCH, Weinheim, **1965**.
- [123]G. I. Fray, R. G. Saxton, *The Chemistry of Cyclooctatetraene and Its Derivatives*, Cambridge University Press, Cambridge, **1978**.
- [124]S. Martin-Santamaria, B. Lavan, H. S. Rzepa, *J. Chem. Soc.: Perkin Trans.*, **2000**, *2*, 1415-1417.
- [125]C. Castro, C. M. Isborn, W. L. Karney, M. Mauksch, P. v. R. Schleyer., *Org. Lett.* **2002**, *4*, 3431-3434.
- [126]C. Castro, Z. Chen, C. S. Wannere, H. Jiao, W. L. Karney, M. Mauksch, P. Ralph, N. J. R. v. Eikema Hommes, P. v. R. Schleyer, *JACS* **2005**, *127*, 2425-2432.
- [127]D. Ajami, K. Hess, F. Köhler, C. Näther, O. Oeckler, A. Simon, C. Yamamoto, Y. Okamoto, R. Herges, *Chem. Eur. J.* **2006**, *12*, 5434-5445.
- [128]J. K. Kang, C. B. Musgrave, *J. Chem. Phys.* **2001**, *115*, 11040-11051.
- [129]a) J. F. M. Oth, H. Röttele, G. Schröder, *Tetrahedron Lett.* **1970**, *1*, 61-66. b) J. F. M. Oth, J. M. Gilles, G. Schröder, *Tetrahedron Lett.* **1970**, *1*, 67-72.
- [130]T. A. Keith, R. F. W. Bader, *Chem. Phys. Lett.* **1993**, *210*, 223-231.
- [131]OooStat 0.4, ein Plugin für OPENOFFICE 2.3.1,  
[http://sourceforge.net/project/showfiles.php?group\\_id=87718&package\\_id=106652](http://sourceforge.net/project/showfiles.php?group_id=87718&package_id=106652).
- [132]OPENOFFICE.ORG 2.3.1, Sun Microsystems Inc., **2007**,  
<http://www.openoffice.org>.
- [133]a) F. Sondheimer, Y. Gaoni, *JACS* **1961**, *83*, 4863. b) I.C. Calder, Y. Gaoni, F. Sondheimer, *JACS* **1968**, *90*, 4946.
- [134]G. Schröder, J. F. M. Oth, *Tetrahedron Lett.* **1966**, *34*, 4083-4088.
- [135]a) C. D. Stevenson, T. L. Kurth, *JACS* **1999**, *121*, 1623. b) C. D. Stevenson, T. L. Kurth, *JACS* **2000**, *122*, 722.

- [136]R. P. Pemberton, C. M. McShane, C. Castro, W. L. Karney, *JACS* **2006**, *128*, 16692-16700.
- [137]S. M. Johnson, I. C. Paul, G. S. D. King, *J. Chem. Soc. B* **1970**, 643-649.
- [138]Andreas Fuchs, *Dissertation an der Christian-Albrechts-Universität zu Kiel - Design und Synthese von Liganden für das Lektin FimH*, Kiel, **2005**.
- [139]R. Herges, A. Dikmans, U. Jana, F. Köhler, P. G. Jones, I. Dix, T. Fricke, B. König, *Eur. J. Org. Chem.*, **2002**, 3004-3014.
- [140]M. N. Braten, C. Castro, R. Herges, F. Köhler, W. L. Karney, *J. Org. Chem.* **2008**, in print, DOI: 10.1021/jo702412d.
- [141]MICROSOFT VISUAL C++ / MICROSOFT VISUAL STUDIO 2005. *Build 8.0.50727.42.*, Microsoft Foundation **2005**.
- [142]G++ 4.1.1, FSF (Free Software Foundation) **2006**, <http://gcc.gnu.org/>.
- [143]*Cygwin: Eine Unix Kompatibilitätserweiterung für Windows. Enthält die typischen GNU-Kommandozeilen-Programme und andere separat versionierte Pakete zum Einsatz unter Windows*, © Red Hat, Inc., <http://www.cygwin.com>.
- [144]THAPI.DLL ist eine Bibliothek des Programmes HYPERCHEM, welche es ermöglicht die Funktionen des Programmes durch ein Programmierinterface der Sprache TCL/TK aufzurufen und Ergebnisse abzufragen.
- [145]TCL/TK *Version 8.04 für Windows*, **1998**, <http://prdownloads.sourceforge.net/tcl/tcl804.exe?download>.
- [146]BOOST::UBLAS ist eine Matrix-Bibliothek für die Programmiersprache C++ und Teil des Boost Projekts, BOOST 1.33.1, **2005**, <http://www.boost.org>.
- [147]*Hyperchem 7.5 Reference Manual, Appendix A*, Hypercube, Inc., Gainesville, Florida **2003**, 623-629.
- [148]GNU DIFFUTILS 2.8.1, Free Software Foundation **2002**, <http://ftp.gnu.org/pub/gnu/diffutils/diffutils-2.8.1.tar.gz>.







## Tabellarischer Lebenslauf:

Name:	Felix Johannes Köhler
Geburtsdatum:	7.4.1973
Geburtsort:	Idar-Oberstein, Deutschland
Eltern:	Christiane, geb. Kühn, Apothekerin Klaus, Oberstleutnant a.D. bei der Bundeswehr
1979-1983	Stockhausengrundschule in Neustadt am Rübenberge
1983-1985	Orientierungsstufe Süd in Neustadt am Rübenberge
1985-1992	Gymnasium in Neustadt am Rübenberge
19.5.1992	Abitur
1.7.92 - 30.6.93	Ableistung des Wehrdienstes
1.10.1993	Anfang des Chemiestudiums an der TU Braunschweig
25.8.1995	Erhalt des Vordiploms
Sept. 95 - Juli 96	ERASMUS Studium in Bordeaux: Synthese und <sup>31</sup> P-NMR Studien über phosphorylierte Zucker
29.3.1998	Diplomprüfung Chemie an der TU Braunschweig
1.4.-29.9.1998	Diplomarbeit an der TU Braunschweig
29.9.1998	Erhalt des Chemie Diploms
1.1.1999-1.1.2001	Stipendium des Fonds der Organischen Chemie
1.1.2001-	Wissenschaftlicher Mitarbeiter im Arbeitskreis Herges, Anfertigung einer Dissertation im Bereich „molecular modelling“ des Faches Organische Chemie, Installation und Wartung von Hard- und Software der Arbeitsgruppe, sowie Betreuung diverser Computerchemiepraktika
11.6.2008 (geplant)	Verteidigung der Doktorarbeit: „Konformations- und Konfigurationshyperflächen - Generierung, Eigenschaften und Darstellung“ an der Universität Kiel



## Ich danke:

Dem AK Herges inklusive aller Ehemaligen, die inzwischen von Süd nach Nord über Deutschland verteilt sind. Insbesondere auch Herr Dr. Antonius Dikmans, der mir immer mit einer erfrischenden Szene im KdW in Erinnerung bleiben wird..

Herrn Dr. Torsten Winkler, sowohl als gutem Kollegen als auch als gutem Freund. Für zahlreiche Grillsessions, für die er das beste Gemüse der Welt besorgt hat, für das gewissenhafte Korrigieren dieser Arbeit, tägliche Fahrten zum Landeshaus und „geekige“ Diskussionen zum Thema Computer.

Jens Walther als größtem Grillmeister aller Zeiten und ausgezeichnetem Koch der lokalen Küche.

Regina Meinlschmidt, die als Institutssekretärin ihren Hofstaat immer resolut, kompetent und fürsorglich verwaltet und sich auch um meine kleinen und großen Probleme liebevoll gekümmert hat.

Herrn Dr. Andreas Fuchs, mit dem ich mir lange Zeit ein „Zimmerchen“ geteilt habe und der mir eine erfolgreiche Einführung in die bayerische Kultur samt ihrer Delikatessen gegeben hat (oder war es die fränkische?). Auch von seinen biologischen Fachkenntnissen (fliegende Pinguine) konnte ich natürlich nur profitieren.

Den nichtwissenschaftlichen Mitarbeitern des Instituts für alle freundlichen Hilfestellungen, die das Leben und Arbeiten einfacher machen. Namentlich: Monika Bänsch (Reinigung), Britta Harbaum und Hannelore Pohl(CTA), Rüdiger Kargoll (Werkstatt), Andreas Wilms (Chemikalienausgabe und Herr der Kfz-Tricks), Katrin Becker und Manuela Krannich (Bibliothek und Adoptivarbeitskreisler) und Nikola Schmitz (ebenfalls Bibliothek), Frau Drieling (ausführliche Einweisung am IR-Gerät).

Herrn Tobias Brich, mit dem ich mich viele Stunden über alle wichtigen und unwichtigen Dinge austauschen konnte.

Meiner Verlobten, Nadine Weiße, die mich immer wieder auf den Boden der Tatsachen geholt hat und als meine große Liebe all meine kleineren und größeren Macken auch in schwierigen Zeiten klaglos erduldet hat, vor allem die größeren. Möge es lange so bleiben.

Meiner Mutter Christiane, meinem Vater Klaus, meiner Tante Roswitha und Onkel Wolfram, die mich finanziell und vor allem moralisch sehr unterstützt haben und mein Wohlergehen nie aus dem Auge verloren haben. Auch um die Rechtschreibkorrektur dieser Arbeit haben sie sich verdient gemacht.

Last, but not least, Frau Prof. Dr. Anke Krüger, die mir als inoffizielle „Zweitdoktormutter“ gerade in der letzten Phase der Arbeit zur Seite gestanden hat mit Aufmunterung, Diskussionen, fachlicher Unterstützung, aber auch dem einen oder anderen Tritt in den Allerwertesten. Ohne sie wäre ich wahrscheinlich immer noch nicht fertig.

Danke, Anke

