Abschlußbericht zum DFN Projekt

Dynamische Moleküle:

Darstellung und Analyse der Dynamik von biologischen Makromolekülen per Internet

http://www.md-simulations.de



Heidelberg, den 31.10.2003

DKFZ-Zentrale Spektroskopie, Im Neuenheimer Feld 280, 69120 Heidelberg, Tel.: 06221-424541; FAX: 06221-422995, Email: {w.vonderlieth,m.frank@dkfz.de}

0 2	ZUSAMMENFASSUNG DER ERZIELTEN ERGEBNISSE	4
1 E	EINLEITUNG	5
1.1	allgemeiner Hintergrund	5
1.2	Bedeutung der Dynamik von molekularen Struktur für biomolekulare Fragestellungen	•
1.3	Die simulierbare Zeitskala	7
1.4	Verfügbarkeit von Molekular Dynamik Simulations-Programmen	7
1.5	Ziele des Projekts	8
	DAS PORTAL ,DYNAMISCHE MOLEKÜLE' AUS DER SICHT DES UTZERS.	g
2.1	Eingangsseite	10
2.2. 2.2. 2.2.	e	10 12 23
2.3	Allgemeines zum Portals 'Dynamische Moleküle' (About)	33
2.4	Informationen zum Portals 'Dynamische Moleküle' (Information)	33
2.5	Depot der benötigten Software und verwandte Anwendungen (Link-Liste)	34
	TECHNISCHE REALISATION: INTERNE UND EXTERNE KOMMUNII 36	CATION
3.1	Grundlegende Konzepte : Auswahl von Soft- und Hardwarekomponenten	36
3.2. 3.2. 3.2. 3.2.	2 Interner Datenfluss	35 38 39 lationen.
3.2.	4 Conformational Analysis Tools (CAT)	41
3.3. 3.3. 3.3.		42 42 43
4 E	BEKANNTMACHUNG DES PORTAL UND BENUTZERRESONANZ	45
4.1	Publikationen	45
4.2	Vorträge und Poster auf wissenschaftlichen Tagungen	40
4.3	Seminare und Fortbildungen	47

4.4	Nutzungsstatistiken	47
5	AUSBLICK	50
6	VERWENDETE TECHNOLOGIEN	52
6.1	Apache	52
6.2	Javascript	52
6.3	Linux	52
6.4	Mac OS X	53
6.5	MySQL	53
6.6	РНР	53

0 Zusammenfassung der erzielten Ergebnisse

Das Ziel des Pilot-Projektes, Lehrern, Studenten und Wissenschaftlern, die sich für die dynamischen Eigenschaften von Biomolekülen interessieren, per Internet einen effizienten und einfach zu handhabenden Zugang zu modernen Computermethoden - insbesondere Moleküldynamik (MD) Simulationen - zu eröffnen, wurde in vollem Umfang erreicht. Das im Rahmen dieses Pilot-Projektes geschaffene Portal ,Dynamische Moleküle' bietet damit erstmalig die Möglichkeit, dass sowohl die Auswahl der zu untersuchenden Substanz, die Einstellung der Simulationsbedingungen als auch die Analyse der Ergebnisse über das Internet mittels Standard-Browser und wenigen, kostenlos verfügbaren, Plug-ins zur Visualisierung realisiert werden kann.

Verschiedene Arten der Benutzerinteraktion werden angeboten: in einem "Anfänger-Modus", der komplett interaktiv abläuft, werden einige klar definierte, vom Benutzer nur begrenzt veränderbare, Instruktionsanweisungen für die Simulationen verfügbar gemacht. Die zur Verfügung gestellten Moleküle – kleine organische Moleküle, Peptide, Kohlenhydrate, DNA – stellen einen repräsentativen Querschnitt der bioorganischen Chemie dar. Im Vordergrund des "Anfänger-Modus" stehen didaktische Ziele, es soll ein allgemeiner Eindruck der Beweglichkeit von Biomolekülen vermittelt werden. Tutorials, in denen anhand typischer Beispiele sowohl die wissenschaftliche Fragestellung und als auch das konkrete Vorgehen dargestellt wird, unterstützen die didaktischen Ziele des Anfänger-Modus. Schüler der Oberstufe, Lehrer und Studenten sind die primäre Zielgruppe.

Der "Experten-Modus" richtet sich an Wissenschaftler, deren Intention es ist, bestimmte wissenschaftliche Fragestellungen mit den angebotenen Werkzeugen zu lösen. Im Rahmen dieses Pilotprojektes wurde exemplarisch für die Substanzklasse der Kohlenhydrate ein allgemeiner Zugang geschaffen, der es erlaubt, alle für diese Klasse von Molekülen relevanten strukturellen Fragestellungen mittels MD Simulationen zu bearbeiten. Der Benutzer hat die Möglichkeit beliebige räumliche Zucker-Strukturen zu konstruieren, die konkreten Simulationsbedingungen frei zu wählen und mitels einer Skriptsprache eine detaillierte Analyse der Dynamik aller interessierenden geometrischen Kenngrößen durchzuführen. Für die Substanzklasse der Kohlenhydrate ist eine automatische Analyse aller üblicherweise analysierten räumlichen Eigenschaften implementiert. Die Ergebnisse der Berechnungen sind in einer Datenbank in Form von Daten und Graphiken verfügbar, auf die alle an einem Projekt beteiligten Personen gleichzeitig Zugriff haben. Dies hat sich für die detaillierte Diskussion der Ergebnisse mit Benutzern/Projektpartnern als ein sehr effizientes Werkzeug erwiesen für die Diskussion von Ergebnissen zwischen räumlich getrennt arbeitenden Wissenschaftlern.

Angesichts der großen Anforderungen an benötigten Rechenzeiten zur Durchführung der Simulationen ist die Nutzung dieser Pilotinstallation sinnvoll für Moleküle bis zu einer maximalen Größe von etwa 1000 Atomen. Simulationen von großen Proteinen mit einer wässrigen Umgebung würden die derzeit verfügbare Hardware für Tage, wenn nicht Wochen blockieren. Derartige Simulationen wären sinnvoller weise im Rahmen von GRID-Computing Projekten angesiedelt. Die Erweiterung des Portals "Dynamische Moleküle" im Hinblick auf derartige Anwendungen ist ein nächster sinnvoller Schritt.

Die Benutzerführung, die geeignete Auswahl der Anwendungen und Tutorien sowie der Einsatz der Analyse-Strategien im Rahmen wissenschaftlicher Projekte wurden in verschiedenen Kooperationen und Demonstrationen getestet und optimiert. Das Portal wurde auf verschiedenen nationalen und internationalen Konferenzen vorgestellt und erzielte dort rege Aufmerksamkeit. Die Molecular Modeling und Graphis Socienty – Deutschsprachige Sektion hat das Portal mit dem Web-Award 2003 ausgezeichnet. Bisher ist eine Publikation im Journal of Molecular Modeling erschienen, in der die grundlegenden Ideen und Realisierung des Projektes beschrieben werden. Weitere Publikationen, insbesondere die Beschreibung eines kompletten wissenschaftlichen Projektes, das auf der Basis von 'Dynamische Moleküle' realisiert wurde, sind in Vorbereitung. Seit Ende März 2003 wurden mehr als 25.000 Web-Seiten von Nutzern aus über 20 Ländern abgerufen. Zwei Hochschullehrer haben angekündigt, das Portal in Ihren Untersicht/Vorlesung einzubauen.

1 Einleitung

1.1 Allgemeiner Hintergrund

Das Internet hat die Art und Weise der wissenschaftlichen Kommunikation wesentlich verändert: Primärdaten wie z.B. räumliche Strukturen von Molekülen können innerhalb von wenigen Sekunden von weltweit verteilten Servern heruntergeladen und visualisiert werden. Die elektronische Verfügbarkeit von wissenschaftlichen Publikationen erspart in vielen Fällen den Gang in die Bibliothek und die kostenlos verfügbaren, sehr effizienten Suchmaschinen erlauben einen schnellen Zugriff auf Texte, Daten und Informationen, die sich auf den Homepages der jeweiligen Forschergruppen befinden. Insgesamt sind aufgrund dieser technischen Entwicklung eine große Vielfalt von Daten und Informationen auf einfachem Wege zugänglich geworden. Gleichzeitig nimmt in der biomedizinischen Forschung die Tendenz zu, komplexe Fragestellungen anzugehen, die sich oft nur durch eine effiziente Zusammenarbeit von Arbeitsgruppen mit unterschiedlichen wissenschaftlichen Schwerpunkten erfolgreich bearbeiten lassen. Da die benötigten Experten häufig räumlich weit voreinander entfernt lokalisiert sind, bietet das Internet günstige Möglichkeiten, auf vielfältige Art und Weise zu kommunizieren. Für eine effiziente Analyse der erzeugten Primärdaten und die Diskussion der Ergebnisse ist es notwendig, dass sowohl ein schneller Zugriff auf alle relevanten Daten erfolgen kann, als auch einheitliche Formate und Schnittstellen vorhanden sind.

Vor diesem Hintergrund haben wir das Projekt 'Dynamische Moleküle' gestartet mit dem Ziel, die wissenschaftlichen Möglichkeiten, die sich durch die Verwendungen molekulardynamischer Simulationen für die biomolekulare Forschung erschließen, einem breiteren Publikum bekannt zu machen. Dabei wollten wir uns nicht auf die reine Darstellung der Möglichkeiten mit im voraus berechneten Daten beschränken, sondern zeigen, dass es unter Verwendung von Standard-Software möglich ist, einen realen Zugriff auf die Primärdaten von Moleküldynamik-Simulationen zu ermöglichen. Dabei wurde besonderes Augenmerk auf die Möglichkeiten der interaktiven Auswertung der Primärdaten gelegt. Mit dem Portal 'Dynamische Moleküle' wurde damit weltweit erstmalig der internet-basierte Zugang zu einer hochkomplexen Simulationsmethode im Bereich der computer-basierten Analyse der Dynamik von Molekülen realisiert. Gleichzeitig sollten unterschiedliche Methoden der projektbezogenen wissenschaftlichen Kommunikation implementiert, untersucht und in der Praxis getestet werden.

Die Molecular Modeling Gruppe der Zentralen Spektroskopie verfügt über eine mehr als zehnjährige Erfahrung im Bereich der Durchführung von wissenschaftlichen Projekten im Bereich der biomolekularen Forschung unter Verwendung von Moleküldynamik-Simulationen. Weiterhin wurde in den letzten Jahren versucht, alle in der Arbeitsgruppe entwickelten Anwendungen und Datenbanken über das Internet verfügbar zu machen, so dass eine großer Erfahrungsschatz in der praktischen Realisierung entsprechender Projekte besteht.

Der vorliegende Abschlußbericht beginnt mit einer Einführung in den wissenschaftlichen Hintergrund der Moleküldynamik, in dem die Bedeutung der Methode für biomedizinische Fragestellungen dargestellt wird. Der zweite Teil stellt die Strukturierung des Portals 'Dynamische Moleküle' aus der Sicht des Benutzers dar. Der dritte Teil beschreibt die technische Realisierung der einzelnen Komponenten. Der abschließende vierte Teil fasst die gewonnenen Erkenntnisse zusammen und versucht einen Ausblick für die weitere Entwicklung des Portals zu geben. Im Anhang befinden sich diverse Dokumente, die unsere Anstrengungen dokumentieren, das Portal bekannt zu machen.

1.2 Bedeutung der Dynamik von molekularen Struktur für biomolekulare Fragestellungen

Alle Moleküle - Gase wie Sauerstoff oder Methan genauso wie Arzneimittel und große Eiweißstoffe (Proteine) - sind keine starren Gebilde, sondern sie bewegen sich ständig in der Luft, in einer Flüssigkeit oder sogar im Festkörper. Das Leben selbst – also z.B. die Teilung von Zellen – wäre nicht denkbar, wenn biologische Makromoleküle, welche die Erbinformation enthalten, ihre räumliche Gestalt nicht ändern könnten. Moleküle können sich voneinander trennen und wiederfinden. Molekulare Erkennung von zwei oder mehreren Molekülen sind verantwortlich für die Steuerung vieler biochemischer Abläufe im Stoffwechsel. Der Transport von Ionen durch Zellmembranen bewirkt die Übertragung von Nervensignalen.

Die experimentelle Untersuchung der Dynamik von Molekülen ist aufwendig. Die Standardmethoden zur Untersuchung der räumlichen Struktur von Molekülen – Röntgenstrukturanalyse und NMR-Methoden – liefern zumeist nur ein statisches Bild. Nur indirekt kann auf die Beweglichkeit bestimmter Teile eines Moleküls zurück geschlossen werden. Hier können nun Computerprogramme helfen, die in der Lage sind, die Bewegungen der einzelnen Atome innerhalb eines Moleküls und ihrer Bedeutung für die Gesamtgestalt des Moleküls zu simulieren.

Wissenschaftler haben Rechenprogramme entwickelt, die in der Lage sind, die Beweglichkeit und Dynamik von Molekülen zu simulieren; die einzelnen Momentaufnahmen der Bewegung können als Film veranschaulicht werden. Die erste Moleküldynamik-Simulation eines kleines Protein ohne Berücksichtigung der umgebenden Wassermoleküle wurde Ende der 70er Jahre beschrieben. Während der letzten 25 Jahre war die Moleküldynamik ein sich rasch entwickelndes Feld der biomolekularen Forschung und hat sich in der Zwischenzeit etabliert als wichtige Methode zur Untersuchung und Beschreibung des dynamischen Verhaltens von Biomolekülen, die insbesondere als Ergänzung und Kombination mit experimentellen Techniken eine hohe Aussagekraft bewiesen hat.

Da die Beweglichkeit eines jeden Atoms simuliert werden muss, handelt es sich bei der MD um eine Rechenzeit intensive Simulationsmethode. Wie alle computerorientierten Zweige der Wissenschaft, hat die Moleküldynamik stark profitiert von der enormen Entwicklung der Hardware. Simulationen werden an

zunehmend größeren molekularen Systemen und unter immer realistischeren Randbedingungen wie Einbeziehung der umgebenden Wassermoleküle durchgeführt. Simulationen, die vor wenigen Jahren nur mittels großer und teurer Supercomputer realisiert werden konnten, können nun mit Standard-Hardware auf Personalcomputern durchgeführt werden. Aufgrund der durch diese technische Entwicklung möglichen längeren Simulationszeiten sowie der Verbesserung bei den Algorithmen und Simulationsprotokollen können heute Eigenschaften von Biomolekülen mit einer Genauigkeit berechnet werden, die vor zehn Jahren nicht erreichbar schien. Simulationen des Transports von Molekülen durch transmembrane Kanäle sind bereits Gegenstand der aktuellen Spitzenforschung.

1.3 Die simulierbare Zeitskala

Trotz dieses enormen Fortschritts bei der Leistungsfähigkeit von Computern, können moleklare Bewegungen, die sich im Bereich von Sekunden abspielen, wie etwa die Faltung von großen Proteinen, noch nicht simuliert werden. Die Moleküldynamik berechnet die Bewegung jedes einzelnen Atoms. Die schnellste Bewegung in einem Molekül – die Kohlenstoff -Wasserstoff Streckschwingung – ist extrem schnell und vollzieht sich im Bereich von 10⁻¹⁴ Sekunden. Damit diese Schwingung durch die Simulation korrekt beschrieben wird, benötigt man mindest 10 Momentaufnahmen, was bedeutet, dass der intrinsische Zeitschritt einer Simulation 10⁻¹⁵ Sekunden (Femtosekunde) beträgt. Das heißt aber auch, dass 10¹⁵ Simulationsschritte berechnet werden müssen, um den Sekundenbereich zu erreichen. Bedenkt man fernerhin, dass eine Simulation eines Proteins unter realistischen Bedingungen die Einbeziehung der wässrigen Umgebung erfordert, so muss man von molekularen Ensembles ausgehen, die einige zehn- bis hunderttausend Atome umfassen. Diese Betrachtung macht die enorme Menge an erforderlicher Rechenleistung aber auch an benötigtem Speicherplatz deutlich. Die längsten heute bekannten Simulationen dringen in den Bereich von einer millionsten Sekunde (Microsekunde 10⁻⁶) vor. Für kurze Modell-Peptide, die sich in dieser Zeitdomäne reorganisieren, konnte kürzlich gezeigt werden, dass MD Simulationen in der Tat geeignet sind, den Prozess der Faltung mit hinreichender Genauigkeit zu berechnen.

Da die Art und Weise, wie Proteine falten eines der großen ungelösten wissenschaftlichen Probleme der Biologie ist und es bisher keine verlässlichen Methoden gibt, das Faltungsmuster eines Proteins aufgrund seiner Primärstruktur vorherzusagen, ist zu erwarten, dass in Zukunft die Moleküldynamik intensiv zur Untersuchung dieser Fragestellung eingesetzt werden wird.

Ein weiteres wichtiges Einsatzgebiet für MD Simulationen ist die Verfeinerung der mittels der Röntgenstrukturanalyse und NMR Spektroskopie erhaltenen räumlichen Strukturen von Molekülen.

1.4 Verfügbarkeit von MD Simulations-Programmen

Verschiedene MD Simulations-Programme stehen dem akademischen Benutzer zur Verfügung und werden weltweit intensiv genutzt. Die traditionell verwendeten Programme CHARMM, AMBER and GROMOS, wurden während der 80er Jahre und entwickelt und werden kontinuierlich gepflegt. TINKER, NAMD and GROMACS (GROningen MAchine for Chemical Simulations) sind für akademische Nutzer frei verfügbare Programme, die in den 90er entwickelt wurden. NAMD und GROMACS

basieren auf einem parallelen, objekt-orientierten Design, das speziell entwickelt wurde für die Simulation von großen biomolekularen Systemen. NAMD skaliert noch gut für 20 bis 30 Knoten, die über eine Fast Ethernet Verbindung miteinander kommunizieren. GROMAS wurde besonders optimiert für die Berechnung der nichtbindenden Wechselwirkungen, die normalerweise die rechenintensivsten Schritte während einer MD Simulation sind. TINKER ist eine Sammlung von Software-Tools für molekular-mechanische und molekular-dynamische Berechnungen, die auch einige Funktionen für Biomoleküle enthält.

Das hauptsächliche Ziel aller erwähnten MD Programme ist, ein repräsentatives Ensemble von Strukturen zu berechnen unter der Annahme von bestimmten Randbedingungen wie Temperatur, Druck und Simulationsdauer. Auch experimentell bestimmte Nebenbedingungen wie etwa die Nachbarschaft von bestimmten Atomen könnten direkt als Randbedingungen in die Simulation einfließen. Darüber hinaus bieten einige Programme auch diverse Funktionen zur Analyse der entstandenen Datenmengen an. Allerdings wird keine oder nur eine sehr einfache graphische Unterstützung angeboten. Für alle Programme gibt es eine mehr oder weniger ausführliche Beschreibung, wie sie zu bedienen sind. Allerdings ist für einen neuen oder gelegentlichen Nutzer die zu überspringende Einarbeitungshürde recht hoch, bevor er die gewünschte Simulation auch tatsächlich starten kann. Aus diesem Grund benutzt zumeist nur eine begrenzte Anzahl von hochqualifizierten und spezialisierten Wissenschaftlern aus dem Bereich des Biocomputing oder der Bioinformatik das breite Anwendungsspektrum tatsächlich, das MD Simulationen für biomolekulare Fragestellungen anbieten.

1.5 Ziel des Projekts

Das Ziel des Projekts 'Dynamische Moleküle' ist es, einen allgemeinen Zugang zu den Techniken der Moleküldynamik Simulationen für einen breiteren Benutzerkreis zu eröffnen. Angesprochen werden sollen Studenten der unteren Semester, Lehrer und Wissenschaftler außerhalb der Bioinformatik. Es ist offensichtlich, dass bei der Komplexität der angewendeten Techniken in jeden Fall ein prinzipielles Verständnis der den Rechenverfahren zugrunde liegenden physikalischen Annahmen notwendig ist, damit die Auswahl der zu untersuchenden Moleküle, die konkrete Wahl Bedingungen einer Simulation und die Länge einer Simulation sinnvoll gewählt werden. Auch ist ein gutes allgemeines Verständnis der verwendeten Methoden hilfreich für die Interpretation der erzielten Ergebnisse. Aus den genannten Gründen werden verschiedene Arten der Benutzerinteraktion angeboten.

Im "Anfänger-Modus", der komplett interaktiv abläuft, werden einige klar definierte, vom Benutzer nur begrenzt veränderbare, Instruktionsanweisungen für die Simulationen verfügbar gemacht. Im Vordergrund des "Anfänger-Modus" stehen didaktische Ziele; es soll ein allgemeiner Eindruck der Beweglichkeit von Biomolekülen vermittelt werden. Deshalb werden auch Tutorials von kompletten Anwendungen angeboten. Schüler der Oberstufe, Lehrer und Studenten sind die primäre Zielgruppe.

Der "Experten-Modus" soll ein breiteres Spektrum an wissenschaftlichen Optionen abdecken, und richtet sich wesentlich an Wissenschaftler, die bestimmte Fragestellungen mittels MD Simulationen untersuchen möchten. Aufgrund der Komplexität der Fragestellungen und der notwendigen Länge der Simulationsdauer ist bei den Projekten des 'Experten-Modus' eine interaktive Nutzung nicht sinnvoll.

Um eine sinnvolle Nutzung des Portals zu fördern werden unterschiedliche Ebenen der Unterstützung sowohl für neue als auch geübte Nutzer angeboten. Allgemeine virtuelle Workshops helfen bei der Einführung in das Thema und zur Erklärung der grundlegenden physikalischen Prinzipien. Spezifische Workshops zeigen den ungeübten Benutzern, wie eine Simulation aufgesetzt wird und wie die entstehenden Rohdaten analysiert werden können.

Das Projekt hatte auch zum Ziel, moderne Methoden der Biochemie über das Internet anzubieten und dabei weitere Zielgruppen – Schüler, Lehrer, Studenten - die Methode der Moleküldynamik näher zu bringen und dabei gleichzeitig die Vorzüge des Internets als wissenschaftliches Kommunikations- und Forschungsmedium zu demonstrieren.

2 Das Portal ,Dynamische Moleküle' aus der Sicht des Benutzers

In diesem Kapitel wird die Strukturierung des Portals "Dynamische Moleküle" aus der Sicht des Benutzers dargestellt. Im Folgenden werden die einzelnen Funktionen und Informationsangebote vorgestellt. Im Mittelpunkt steht dabei die Beschreibung des Simulationsmanagers, der den Kern der Anwendung bildet.

In dem ursprünglichen Antrag war vorgesehen, dass der Experten-Modus auf die Untersuchung von Kohlenhydraten beschränkt wird und deshalb die Verwaltung der Daten von der bereits existierenden SWEET-DB übernommen werden sollte. In Hinblick jedoch auf die beabsichtigte Erweiterung des Portals für weitere Molekülklassen, wurde davon Abstand genommen und eine Infrastruktur von gekoppelten Datenbank-Tabellen geschaffen – eben den Simulationsmanager (siehe auch Kapitel 3.2.2.) –, die in der Lage ist, weitere Klassen von Molekülen zu verwalten. Diese sehr aufwendigen Arbeiten gehen weit über die im Projektantrag zugesagten Leistungen hinaus und zielen auf einen weiteren Ausbau des Portals.

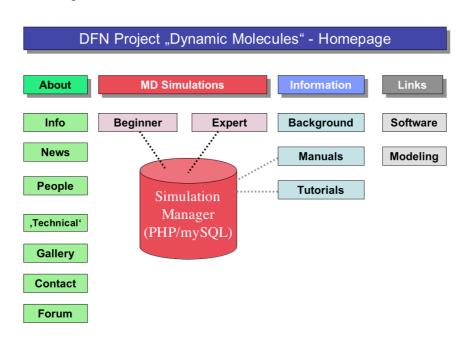


Abbildung 1) Funktionsumfang des Portals "Dynamische Moleküle".



Abbildung 2) Eingangswebseite des Portals 'Dynamische Moleküle'. Durch Aktivierung einer der vier Hauptfunktionen 'About', 'Information', 'Simulationsmanager', 'Links' wird der Benutzer zu den jeweiligen Untermenues geführt.

2.1 Eingangsseite

Die Eingangseite vermittelt den schnellen Zugang zu den einzelnen Haupt-komponenten und Informationsseiten des Portals, die in der Kopfzeile des Portals in weißer Schrift angegeben sind. Der Punkt "About" beschreibt Ziele und Ideen des Projekts. Unter "Information" ist der Zugang zu den Workshops, Manuals und den physikalischen Grundlagen der Methode zu finden. Der "Simulationmanager" ist das eigentliche Herz des Portals und wird im Folgenden deshalb auch als erstes beschrieben. Unter "Links" befinden sich Verweise auf verwandte Web-Seiten und interessante frei verfügbare Software.

Neben den Hinweisen auf die Urheberschaft und die Unterstützer des Projektes wird auch ein "Newsticker" auf der Eingangseite angezeigt, der die Benutzer über die neuesten Entwicklungen des Portals hinweist.

2.2 Der Simulationsmanager

Der Simulationsmanager bietet Zugang zu dem mehr pädagogisch orientierten "Beginner-Modus" und dem "Experten-Modus", der es erlaubt, komplette wissenschaftliche Projekte durchzuführen.



Abbildung 3) Hauptseite des "Simulationmanager". Der Benutzer hat die Auswahl zwischen den "Beginner"- und dem "Experten"-Modus.

Bei Aktivierung der Funktion "Simulationsmanager" erscheint zur allgemeinen Orientierung des Benutzers eine schlagwortartige Beschreibung der insgesamt implementierten Funktionalitäten. Diese untergliedern sich in Möglichkeiten, eine Simulation für ein spezielles Molekül aufzusetzen und durchzuführen ("setup") sowie der nach erfolgter Simulation durchzuführenden Analyse der erzeugten Daten ("Analysis") und deren Visualisierung ("Visualisation"), die mittels verschiedener Hilfsprogramme (Plug-ins) erfolgt. Die Funktion "Retrieval" erlaubt die Suche nach bereits durchgeführten Simulationen und daraus erzeugten Graphiken, die in der zugrunde liegenden Datenbank protokolliert wurden.

Der Simulationsmanager selbst besteht aus einer Datenbank-Tabelle, in der die zugelassen Benutzer registriert werden ("User Manager") und weiteren Tabellen in denen alle bereits durchgeführten Simulationen ("Job Manager") protokolliert, die implementierten Skripte zum Starten Simulationen bereitgehalten ("Script Manager") und konstruierte Molekül-Strukturen ("Structure Manager") verwaltet werden.

2.2.1 Beginner Modus

Der "Beginner" Modus steht ohne Einschränkungen allen Benutzern zur Verfügung. Eine gesonderte Registrierung ist nicht erforderlich. Die Eingangsseite zum "Beginner" Modus beschreibt kurz den Zweck (mehr pädagogische Ausrichtung) dieser Anwendung und benennt die benutzerseitigen Voraussetzungen (Browser und Plug-ins) unter denen die volle Funktionalität garantiert werden kann.

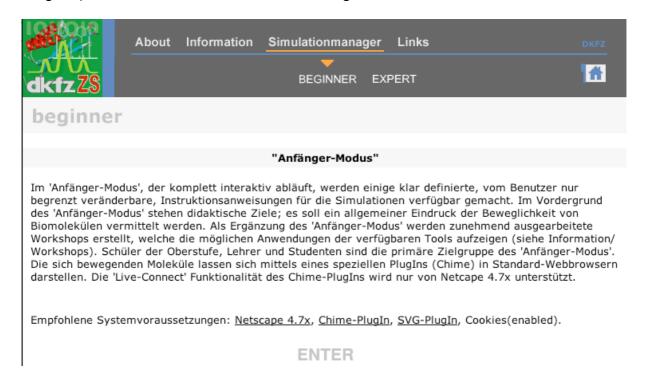


Abbildung 4) Erklärende Eingangsseite zum "Beginner" Modus. Durch Aktivierung der "Enter" Option wird der Benutzer zu dem eigentlichen Simulations-Manager weitergeleitet.

2.2.1.1 Simulationsmanager für den Beginner-Modus

2.2.1.1.1 Auswahl von Molekülen

Die über die Eingangseite des Simulationsmanagers (siehe Abbildung 4) zur Verfügung gestellten Moleküle – kleine organische Moleküle, Peptide, Kohlenhydrate, DNA – stellen einen repräsentativen Querschnitt der bioorganischen Chemie dar. Der Benutzer hat die Möglichkeit eines dieser Moleküle über das Pull-Down Menue zu aktivieren. Die Moleküldarstellung erfolgt sodann mittels des Plugins Chime, das ein kostenlos verfügbares, intensiv verwendetes und intuitiv zu bedienendes Programm zur Visualisierung molekularer Strukturen innerhalb eines Browserfensters ist. Es bietet darüber hinaus für den geübten Benutzer, der bereit ist, sich in die ChimeSkript-Sprache einzuarbeiten, die Möglichkeit der Erzeugung von qualitativ hochwertigen Moleküldarstellungen. Leider ist Chime nicht für alle Browser-Versionen verfügbar. Mit Internet Explorer und Netsacpe 4.7x (Netscape 7 und Mozilla binden Chime ein, allerdings unterstützen die Installationsroutinen von Chime diese Browser nicht, so dass der Benutzer per Hand Dateien in den richtigen Ordner verschieden muss) ist Chime für die am meisten verwendeten Browser verfügbar.

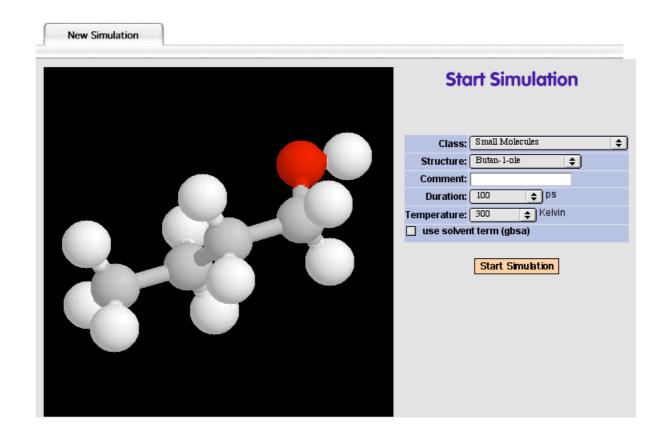


Abbildung 5) Eingangsseite des Simulationsmanagers für dem "Beginner" Modus. Die Moleküldarstellung des Butanols erfolgt mit den *Chime-Plugin*. Eine so genannte "Ball and Stick"-Darstellung wurde gewählt. (Farbkodierung: Wasserstoff: weiß; Kohlenstoff: grau; Sauerstoff: rot)

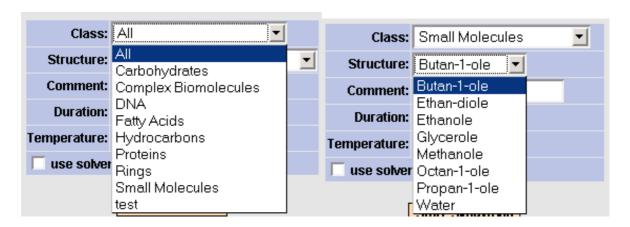


Abbildung 6) Auswahlmenu der für den "Beginner" Modus zur Verfügung stehenden Moleküle. Links: Übersicht der verfügbaren Molekülklassen. Rechts die unter der Klasse "Small Molecules" verfügbaren Moleküle.

2.2.1.1.2 Starten von Simulationen

Nach Auswahl des zu untersuchenden Moleküls kann der Benutzer noch die Simulationstemperatur und die Länge der Simulationsdauer in Picosekunden (10⁻¹² Sekunden) wählen. Die Aktivierung eines Terms zur Beschreibung des Einflusses von Wasser ist nur für Peptide sinnvoll. Ist die Eingabe erfolgt, so kann die

Simulation gestartet werden durch Klicken auf den "Start Simulation" Knopf. Da im Beginner' Modus nur eine begrenzte Anzahl von Molekülen und Simulationsbedingungen auswählbar sind, ist die Anzahl der möglichen Simulationen begrenzt. Um wertvolle Rechenzeit nicht unnütz zu verbrauchen wird, bevor eine neue Simulation gestartet wird, deshalb zunächst die Job-Datenbank abgefragt, ob das zu untersuchende Molekül bei der gewünschten Temperatur und Simulationsdauer bereits zuvor berechnet wurde. Ist dies der Fall, werden die Simulationsdaten direkt aus der Datenbank abgerufen, was zur Folge hat, dass der Benutzer sofort in das ,Display Simulation' Fenster weitergeleitet wird. Wurde das gewünschte Molekül noch nicht unter den eingegebenen Bedingungen simuliert, so werden die entsprechenden Simulations-Skripte und Koordinaten an den im Hintergrund laufenden LINUX Cluster transferiert und dort das Simulationsprogramm gestartet. Je nach Größe der Moleküle, Länge der zu simulierenden Zeitdauer und Auslastung des Clusters kann es dann zu Wartezeiten von einigen Minuten oder auch länger kommen. Der Benutzer hat die Möglichkeit, den aktuellen Bearbeitungszustand jederzeit abzufragen. Da davon ausgegangen werden kann, dass viele der im "Beginner"-Modus verfügbaren Moleküle bereits berechnet und ihre Simulationsdaten in der Datenbank verfügbar sind, werden die Abfragemöglichkeiten des Bearbeitungszustands einer Simulation in im Zusammenhang mit dem "Experten'-Modus diskutiert. Es ist zu erwarten, dass nach und nach von allen Molekülen und bei allen vorgegebenen Temperaturen Simulationen durchgeführt werden, so dass zunehmend häufiger direkt auf die abgespeicherten Daten zurückgegriffen und man deshalb in der Tat von einer Interaktivität des "Beginner"-Modus sprechen kann.

2.2.1.1.3 Wiedergabe von Simulationen

Das "Display Simulation" (Abbildung 6) Fenster erlaubt die interaktive Wiedergabe der Bewegungen der untersuchten Moleküle während einer MD Simulation bei einer bestimmten Temperatur. Das Plug-in Chime hat sich auch für die Wiedergabe der zeitlichen Abfolge von Molekülstrukturen als besonders leistungsfähig erwiesen. So kann während des Abspielens einer Bildsequenz mit der Maus das Molekül beliebig bewegt werden. Auch aufwändige Moleküldarstellungen wie etwa Kalottenmodelle werden praktisch ohne zeitliche Verzögerung dargestellt. Der die Geschwindigkeit des Abspielens bestimmende Schritt ist zumeist das Laden der Datei über das Internet, in der die Molekülkoordinaten enthalten sind. Je nach Größe der Moleküle können diese Dateien relativ groß werden (siehe Tabelle 1). Aus diesem Grund bietet das "Display Simulation" Fenster die Option an, dass nur bestimmte Bereiche oder Bereiche mit größeren Intervallen angefordert werden. Für eine erste, rein visuelle Betrachtung der Beweglichkeit eines Moleküls ist diese Form der Darstellung zumeist ausreichend.

Generell werden für den "Beginner'-Modus unabhängig von der gewählten Simulationsdauer immer 1000 Strukturen gespeichert. Der praktische Hintergrund dieser Festlegung ist, dass die auf dem File-Server abgelegten und über das Netz transferierten Simulationsdaten nicht zu groß werden sollen. Vom wissenschaftlichen Standpunkt aus kann argumentiert werden, dass es je nach Fragestellung es unterschiedliche zeitliche Domänen gibt, in denen sich die zu untersuchenden Phänomene ereignen. Während sich z.B. der Übergang von einer Orientierung in eine andere bei den exozyklischen Hydroxylgruppen von Kohlenhydraten im Bereich von wenigen Picosekunden (ps) abspielen – also eine Simulationsdauer von 100 ps zur Untersuchung dieser Übergänge ausreichend ist – werden Übergänge von einer

Konformation in eine andere bei Disacchariden im Bereich von Nanosekunden beobachtet. Folglich braucht man auch eine wesentliche längere Simulationsdauer, um diese Veränderungen überhaupt beobachten zu können. Andererseits kann man dafür dann das Zeitintervall zwischen den einzelnen Momentaufnahmen, die abgespeichert werden, vergrößern ohne wesentliche Informationen zu verlieren.

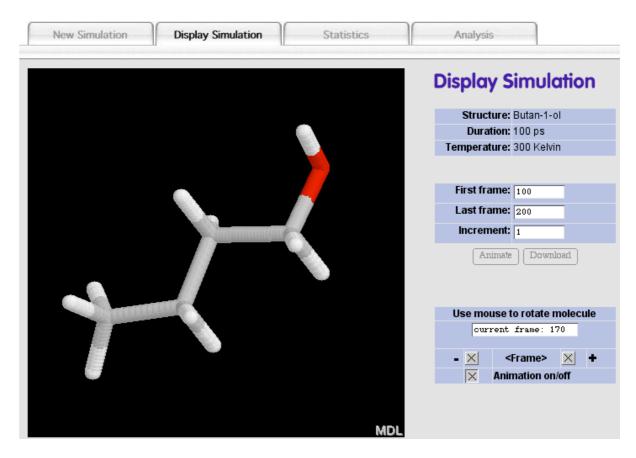


Abbildung 7) Das , *Display Simulation'* Fenster. Angezeigt im Chime-Fenster wird die 170ste Struktur einer 100 ps MD Simulation von Butan-1-ol bei 300 K.

Tabelle 1: Größe der unkomprimierten Archive der Atomkoordinaten (jeweils 1000 Momentaufnahmen im XYZ-Format), die benötigt werden für die Annimation der simulierten Moleküle.

Molekül	Anzahl der Atome	Größe der XYZ Dateien
		[MB]
Methanol	6	0,179
Butan-1-ol	15	0,432
Lactose	42	1,27
Lewis X	71	2,00
GM1-Cer	228	6,39
Peptid	334	9,36
DNA-DS 10 bp	639	17,90
Protein	1022	27,63

2.2.1.1.4 Erstellung von Filmsequenzen

Der ursprüngliche Antrag sah vor, das von den jeweiligen Simulationen auf der Serverseite Filmsequenzen produziert werden, die sodann als Animation direkt in Präsentationen eingebunden werden können. Die erste Realisierung des "Beginner" Modus war auch in der Tat in der Lage, direkt einen Film zu erstellen (Siehe Abbildung 8). Allerdings gab es eine Reihe praktischer Aspekte, die uns bewogen in der finalen Version ein anderes Konzept zu favorisieren.

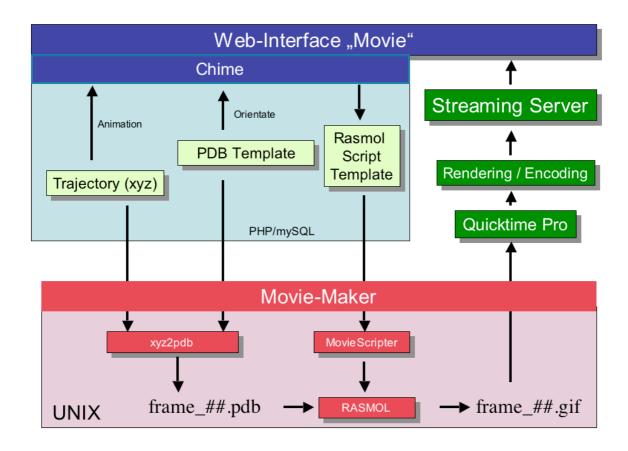


Abbildung 8) Das ursprüngliche 'Make Movie' Konzept.

Da die interaktive Visualisierung von Simulationen mit dem Chime-Plug-in oder dem neuen Java-Applet JMOL eine Reihe von praktischen Vorteilen (Moleküle können während des Abspielens der Simulation per Hand optimal orientiert werden, unterschiedliche Moleküldarstellungen können gewählt werden, je nach Bedarf können bestimmte Molekülteile eingefärbt werden zu ihrer besseren Verfolgung während der Simulation) gegenüber dem Abspielen einer nicht mehr veränderbaren Filmsequenz hat, war schnell klar, dass die Erzeugung und Verwendung von Filmsequenzen eigentlich nur für Präsentationszwecke z.B. innerhalb von Powerpoint Vorführungen sinnvoll ist. Da einerseits die Windows-Version von Powerpoint nur AVI-Filme abspielen kann, die keinen hohen Grad an Komprimierung erlauben ohne dass die Qualität deutlich abnimmt und deshalb immer sehr große Dateien (einige hundert MB) entstehen und da andererseits nicht davon auszugehen ist, dass alle Benutzer des Projektes eine schnelle Anbindung an das Internet haben (Schulen zum Beispiel), entschlossen wir uns, dass die Erzeugung von Filmsequenzen am besten lokal auf den Rechnern der Benutzer erfolgt. Zudem hätte die serverseitige Erzeugung von Videosequenzen bedeutet, dass ein zusätzlicher Rechner für das Film-Rendering beschafft werden müsste, da wir trotz intensiver Suche keine Software zur Visualisierung von Molekülen finden konnten, die im Batch-Modus als Hintergrund-Prozess arbeitet. Alle uns bekannten Produkte brauchen zwingend eine

grafische Ausgabeeinheit, auf der die Abbildungen zunächst gerendert werden, bevor sie als einzelne Bilder abgespeichert werden können.

Die Entscheidung, die Videodatei nicht serverseitig zu erstellen, wurde uns auch dadurch erleichtert, dass verschiedene frei verfügbare Programme (Biodesigner, iMol, VMD) mittlerweile die Option anbieten, direkt Bildsequenzen bzw. Filme von Simulationen zu erzeugen. Mit dem Entwickler des "Biodesigners" und "iMol" Piotr Rotkiewicz findet ein reger Gedankenaustausch statt, der dazu dazu beigetragen hat, dass weitere, von uns gewünschte Funktionen in "iMol" implementiert wurden.

Im Folgenden soll das prinzipielle Vorgehen für das Programm "Biodesigner' beschreiben werden. Im Anhang befindet sich eine ausführliche Benutzeranleitung, wie Animationen mit dem Portal "Dynamische Moleküle" erzeugt und herunter geladen werden können auf den lokalen Rechner und mittels "Biodesigner" daraus ein Film erzeugt werden kann.

2.2.1.1.5 Erstellung von Filmsequenzen mit 'Biodesigner'

Die Rohdaten der Simulation werden durch Aktivieren des "Download" Knopfes im "Display Simulation" Fenster (Abbildung 9) im multiplen PDB-Format auf die lokale Platte herunter geladen. Dabei kann der Benutzer über die Parameter "First frame", "Last Frame" und "Increment" genau den ihn interessierenden Zeitraum mit der gewünschten zeitlichen Auflösung angeben.

Das Programm 'Biodesigner' muss lokal installiert werden. Es kann von der Link-Seite des Projektes herunter geladen werden und installiert sich automatisch für alle Windows Betriebssysteme. Das Laden der Framesequenz (Trajektorie) erfolgt über das File → Open Molecules Menue. Allerdings muss man hierbei darauf achten, dass in den 'Preferences' bei den 'Input/Output Options' der Unterpunkt 'Multiple models as trajectory' zuvor aktiviert wurde. Nach dem erfolgreichen Laden der Trajektorie erscheint das 'Animation Control Panel'. Dieses erlaubt es, die Simulation abzuspielen (unter Animation muss Model aktiviert sein) oder direkt einen AVI File (Angabe des kompletten Dateinamens mit Extension *.avi erforderlich). Danach kann die erzeugte Filmsequenz direkt in eine Powerpoint-Präsentation eingebunden werden oder mittels eines Players (z.B. Windows Media Player) abgespielt werden.

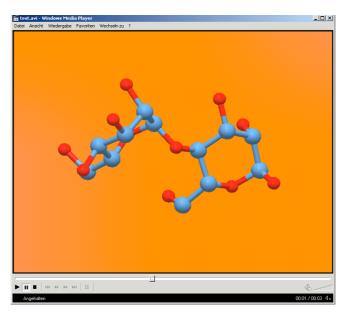


Abbildung 9) Abspielen eines mittels des Programms "Biodesigner" erzeugten AVI-Filmsequenz. Diese Filmsequenz kann auch direkt in Powerpoint-Präsentationen eingebunden werden

2.2.1.1.6 Allgemeine Charakterisierung von MD Simulationen

Das "Simulation Statistics" Fenster (Abbildung 10) gibt einen allgemeinen Überblick über wichtige Kenngrößen einer durchgeführten Simulation. Zur Bewertung, ob eine Simulation stabil abgelaufen ist, werden routinemäßig die Temperatur und die Energie des Gesamtsystems herangezogen. Die mittlere tatsächliche Temperatur sollte nur wenig von der eingestellten abweichen. Minimal und maximal aufgetretene Temperaturwerte sollten nicht mehr als 10% von dem tatsächlichen Mittelwert abweichen und die Steigung sollte einen Wert in der Nähe von Null haben. Ähnliches gilt für den Verlauf der Gesamtenergie und ihrer Komponenten. Für eine detaillierte Analyse von Temperatur und Energie ist die Darstellung des zeitlichen Verlaufs (Abbildung 11) der einzelnen Größen hilfreich.

	Sim	ulation :	Statistic	:S			
Structure: Butan-1-ol	Title: Butan-1-ole, Duration = 100 ps, Temperature = 300 K, Solver				olvent= none	nt = none	
Job Parameters							
Temperature [K]	300		Total time	e [ps]	100		
No of frames	1000		Current time [ps] Finished [%]		100 100		
History frequency [ps]	0.1						
	Mean	Std.Dev.	Min	Max	Slope	Slope(A	
Total Energy [kcal/mole]	28.27	0.26	26.55	29.21	-0.33	0.33	
Pot. Energy [kcal/mole]	16.65	0.35	14.61	17.87	-0.21	0.21	
Kin. Energy [kcal/mole]	11.63	0.23	10.88	12.42	-0.12	0.12	
Temperature [K]	300.01	6.05	280.83	320.58	-3.11	3.11	

Abbildung 10) Das , Simulation Statistics' Fenster gibt einen allgemeinen Überblick über wichtige Kenngrößen zur Bewertung einer durchgeführten Simulation.

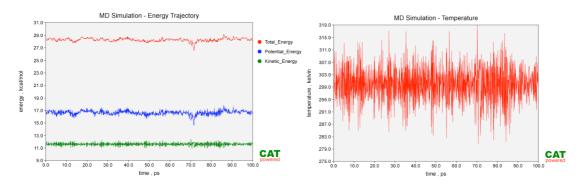


Abbildung 11) Der zeitliche Verlauf von wichtigen Kenngrößen (auch Trajektorien genannt) wie Energie des Gesamtsystems und Temperatur gibt wichtige Hinweise auf die Stabilität der durchgeführten Simulation.

2.2.1.1.7 Auswertung von Simulationen

Das "Analysis" Fenster erlaubt die detaillierte Auswertung des dynamischen Verhaltens aller strukturellen Eigenschaften eines Moleküls. Die eigentliche statistische Auswertung des zeitlichen Verlaufs bestimmter geometrischer Kenngrößen – auch Trajektorien genannt – erfolgt mit dem im Rahmen des Projektes von Dr. Martin Frank weiterentwickelten Programm Conformational Analysis Tools (CAT). Um eine möglichst flexible Interaktion mit CAT zu erreichen, wurde eine XMLbasierte Skript-Sprache entwickelt, die es erlaubt, alle möglichen inneren Koordinaten eines Moleküls (Abstände, Winkel, Torsionswinkel) anzusprechen und bestimmte Abhängigkeiten unter ihnen auszuwerten. Die Definition von Parametern. die analysiert werden sollen erfolgt mittels so genannter "Wave Definitions", die einer bestimmten Syntax unterliegen. Für den "Beginner" Modus wurde ein graphische Schnittstelle geschaffen, die es dem Benutzer erlaubt, ohne Kenntnis der Skript-Sprache die gewünschten Informationen abzufragen. Für viele im "Beginner'-Modus verfügbare Moleküle wurden charakteristische "Wave Definitions" vordefiniert und können über die Option , Get predefined' abgerufen werden. Für das in Abbildung 12 benutze Butan-1-ol sind dies die Torsionswinkel der Schweratome plus der Hydroxylgruppe. Durch Drücken des "Start Analysis' Knopfes wird die Auswertung aller in dem Skript enthaltenen "Wave Definitions" durch das Programm CAT ausgeführt.

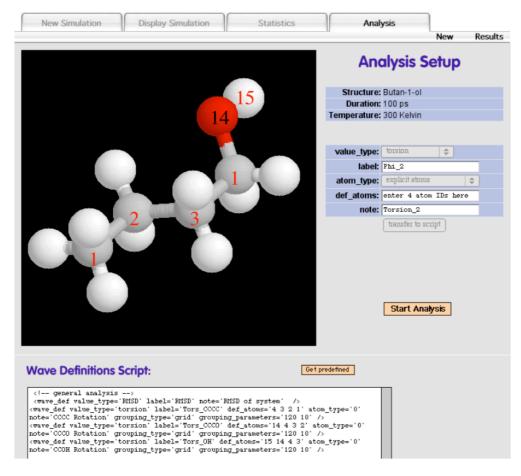


Abbildung 12) Das "Analysis' Fenster erlaubt die detaillierte Auswertung des dynamischen Verhaltens aller strukturellen Eigenschaften eines Moleküls. Aktiviert wurden die für Butan-1-ol vordefinierten "Wave Definitions" durch Drücken des Knopfes "Get predefined". Die Atomnummern der einzelnen Atome sind in der Moleküldarstellung angezeigt.

2.2.1.1.8 Liste der Resultate

Die serverseitige Auswertung erfolgt in aller Regel sehr schnell, so dass schon nach wenigen Sekunden die "Results List" Seite erscheint. In einer tabellarischen Form werden alle bis zu diesem Zeitpunkt ausgeführten Analysen aufgelistet. Loggt sich ein Benutzer zu einem späteren Zeitpunkt über den gleichen Rechner (Identifizierung von Beginner-Usern erfolgt über Cookies) wieder ein, so hat er nach wie vor Zugriff auf die bereits ausgeführten Analysen. Die Liste der Resultate erlaubt es drei unterschiedliche Arten von Daten anzusehen bzw. herunterzuladen: Grafiken (Objekt-Typ: svg), Originaldaten (Objekt-Typ: wave) und charakteristische räumliche Strukturen (Objekt-Typ: pdb).

Das Programm CAT erzeugt automatisch bei der Ausführung der Skripte graphische Darstellungen im SVG-Format (Scalable Vector Graphic), die direkt in einem Browser-Fenster mittels des SVG Plug-ins dargestellt werden können. Durch Drücken des entsprechenden "Show" Knopfes werden die graphischen Darstellungen der Trajektorien aktiviert. In Abbildung 14 ist die Trajektorie der mittleren quadratischen Abweichung (RMSD) aller Atome von Butan-1-ol während der Simulation dargestellt. Die statistischen Kenngrößen der ausgewerteten Daten (wave) können ebenso angezeigt werden wie die räumlichen Strukturen (pdb) von charakteristischen Konformationen, die während der Simulation aufgetreten sind.

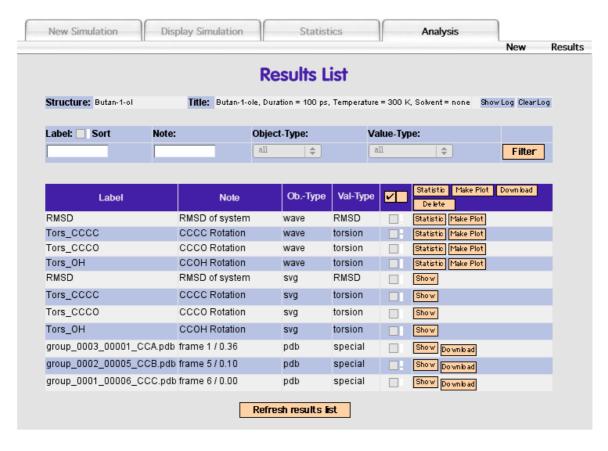


Abbildung 13) Das "Results List' Fenster zeigt alle bereits ausgewerteten Trajektorien einer bestimmten Simulation. In der Spalte "Label" wird der Name einer Trajektorie so wie er in der jeweiligen "Wave Definition" angegeben wurde, angezeigt. In der Spalte "Note" ist ein zusätzlicher Kommentar abgelegt. Die Spalte Object Type (Ob.-Type) beschreibt die Art der Daten. "wave", steht für die Rohdaten, "svg" für bereits erzeugte Grafiken und "pdb" für die Koordinaten der molekularen Strukturen. Alle orange unterlegten Buttons erlauben dem Benutzer eine Ausführung der jeweiligen Funktion.

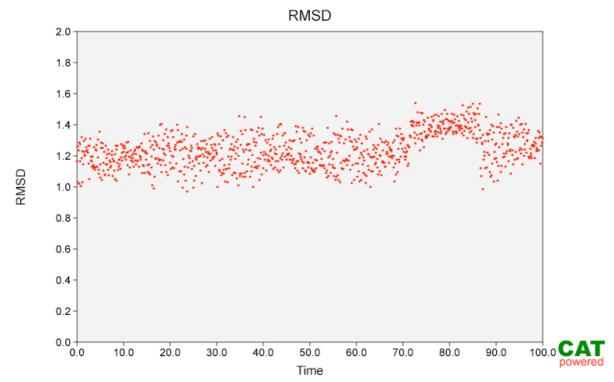


Abbildung 14) Trajektorie der mittleren quadratischen Abweichung aller Atome von Butan-1-ol

2.2.1.1.9 Erzeugung von speziellen Grafiken

Ist die gewünschte Darstellung noch nicht vorhanden – man will z.B. mehrere Trajektorien in einem Bild haben – so kann man auf die entsprechenden Original-Daten (Objekt-Typ ,wave') zurückgreifen, in dem man diese über die Checkboxen aktiviert und den ,Make Plot' Knopf in der Kopfzeile aktiviert (Siehe Abbildung 15) .

Label	Note	ObType	Val-Type		Statistic Make Plot Down load De lete
RMSD	RMSD of system	wave	RMSD		Statistic Make Plot
Tors_CCCC	CCCC Rotation	wave	torsion	☑	Statistic Make Plot
Tors_CCCO	CCCO Rotation	wave	torsion	\checkmark	Statistic Make Plot
Tors_OH	CCOH Rotation	wave	torsion	$\overline{\checkmark}$	Statistic Make Plot

Abbildung 15) Erstellung einer neuen Graphik in der drei Trajektorien gleichzeitig geplottet werden.

In dem nachfolgenden Abfragemenue hat der Benutzer die Möglichkeit, die Art der Darstellung der Trajektorien (Punkte, Linien, Histogramm) auszuwählen und die Beschriftung der Achsen sowie den Titel einzugeben (Abbildung 16). Die neu erstellte Darstellung erscheint sodann in der Resultateliste und kann von dort mittels des üblichen 'Show'-Knopfes aktiviert werden. In Abbildung 17 sind die in Abbildung 15 drei ausgewählten Trajektorien von Butan-1-ol dargestellt. Man erkennt deutlich, dass die ausgewählten inneren Koordinaten sich in unterschiedlichen Zeitdomänen bewegen. Während die beiden Torsionswinkel der Schweratomkette des Butan-1-ol nur innerhalb einer Konformation fluktuieren, zeigt die Hydroxylgruppe viele Übergänge zwischen den drei hauptsächlichen Konformationen bei –60, +60 und 180 Grad und bevölkert darüber hinaus auch noch den sonstigen konformationellen Raum. Dies ist besonders gut sichtbar bei einer Darstellung der Populationen der Torsionswinkel in Form eines Histogramms (siehe Abbildung 18). Aus diesen Daten

kann geschlossen werden, dass die Simulationslänge von 100 ps bei einer Temperatur von 300 K ausreichend ist, um die Beweglichkeit von Hydroxylgruppen zu untersuchen. Sie ist jedoch nicht ausreichend für die Analyse der Beweglichkeit der Schweratomkette von Butan-1-ol, da keine Übergänge zu anderen Konformationen erfolgen. Um die Beweglichkeit dieser beiden inneren Koordinaten zu untersuchen muss entweder die Simulationszeit stark verlängert werden oder aber die Simulationstemperatur muss erhöht werden, so dass die energetische Barriere zwischen den einzelnen Konformationen häufiger übersprungen werden kann.

Structure: Butan	-1-ol Title:	Butan-1-ole, Duration	= 100 ps, Temperat	ure = 300 K, Solvent = none	
Enter paramete	rs of your plot:				
Plot type:	Trajectory \$		Plot style:	Dots 💠	
Label:	Butan-1-ol		Label (x):	Zeit	
Title:	Beweglichkieten im B	utan-1-ol	Label (y):	Torsionswinkel	
		Sul	omit		

Abbildung 16) Auswahlmöglichkeiten bei der Erstellung neuer Grafiken.

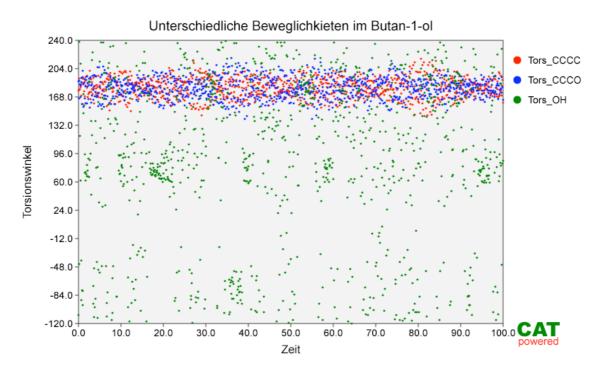


Abbildung 17) Interaktiv erstellte Graphik mit drei Trajektorien aus den Originaldaten

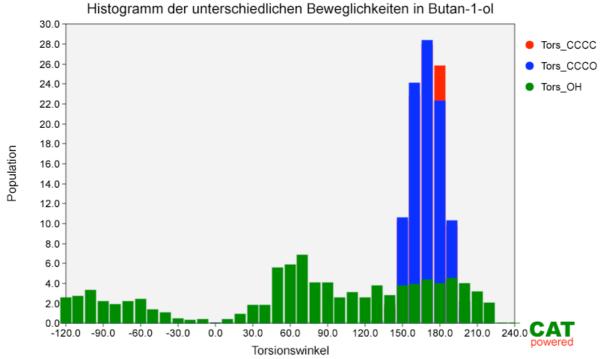


Abbildung 18) Histogramm der Population einzelner Torsionswinkel-Bereiche für die drei Trajektorien, die in Abbildung 15 definiert wurden.

2.2.2 Experten-Modus

2.2.2.1 Zugang zum Experten-Modus

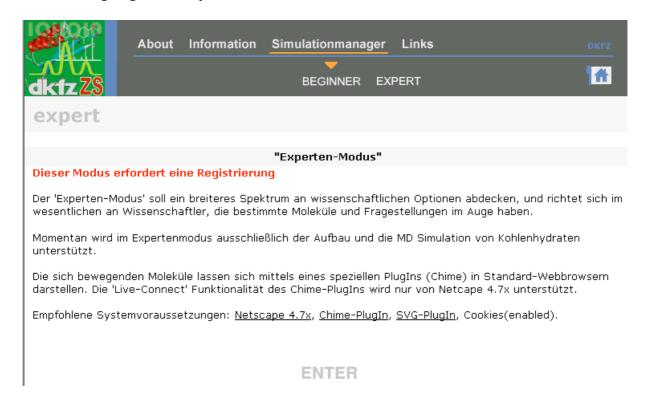


Abbildung 19) Eingangsseite des Experten-Modus

Für die Benutzung des "Experten'-Modus ist eine Registrierung als Nutzer erforderlich. Die Erteilung einer Benutzerkennung erfolgt auf Anfrage. Bisher sind 30 Benutzer als solche eingetragen. Der Zugang zu den Daten ist passwortgeschützt. So kann garantiert werden, dass alle Simulationen und Auswertungen eines Wissenschaftlers vor dem Zugriff durch andere geschützt ist. Dies ist eine wichtige Anforderung an den "Experten'-Modus, da nur so Wissenschaftler davon überzeugt werden können, dass aktuelle Forschungsprojekte nicht ausgespäht werden können. Arbeiten mehrere Wissenschaftler gleichzeitig an einem Projekt, so können auch Kennungen für das jeweilige Projekt vergeben werden, so dass eine klar definierte Gruppe von Wissenschaftlern jederzeit Zugriff auf alle bisher erzeugten Daten hat. Diese Option hat sich als sehr hilfreich und effizient erwiesen und wird in zwei Projekten intensiv genutzt.

Name DFN
Password •••••••

Abbildung 20) Das Login-Fenster für den Experten Modus.

2.2.2.2 Hauptmenus des Experten-Modus

Zum einfachen, übersichtlichen Zugriff auf bereits durchgeführten Simulationen und deren Analysen sowie zum Starten von neuen Berechnungen, wurde eine klare Trennung zwischen der Verwaltung von Jobs, den molekularen Strukturen und den zu einer Struktur gehörenden Auswertungen vorgenommen.

2.2.2.3 Job Liste

Die Verwaltung aller bereits durchgeführten Simulationen eines Benutzers erfolgt über das "Jobs' Menue. Angesprochen werden können die einzelnen Berechnungen durch eine vom Benutzer frei wählbare Job-ID. Die Suchfunktion erlaubt die Auswahl bestimmter Simulationen aufgrund der Job-ID, dem Titel einer Berechnung und ihrem Bearbeitungszustand.

Jobs	Structures	Results	Preferences
		Job List	
Job-ID:	Structure:	Status:	Show ra
		all \$	Search 1 :
Job-ID	Title	started a	rt Status Newsimu
dfn_1	O-GlyKan Core 4,500K,1000ps	2003-10-17 14:0	5:44 waiting
test1	adman-14-adman,300K,1000ps	2003-03-27 22:5	7:11 finished Display
LAC0011	300K simulation	2003-03-21 19:1	5:47 finished Display

Abbildung 21) Verwaltung von Simulationen über die Joblist.

Durch Aktivierung des ,New Simulation' Knopfes können neue Berechnungen gestartet werden. Drücken des ,Display' Knopfes bringt den Benutzer direkt in den Auswertemodus.

2.2.2.3.1 Konstruktion von Molekülen

Der "Experten"-Modus erlaubt - im Gegensatz zu "Beginner"-Modus, wo nur eine charakteristische Auswahl von Molekülen zur Verfügung steht - die Untersuchung von Strukturen, die durch den Benutzer selbst konstruiert oder eingelesen werden können. Als Pilotanwendung wurde die Konstruktion von beliebig komplizierten Oligosacchariden und zusätzlich von linearen Peptiden realisiert. Für diese beiden Substanzklassen konnte auch die Zuordnung zu dem jeweiligen Kraftfeld (MM3 für Oligosaccharide und Amber/Charmm für Peptide) soweit automatisiert werden, dass keine zusätzliche Benutzerinteraktion mehr notwendig ist. Die richtige Zuordnung zum Kraftfeld und das Vorhandensein aller benötigten Parameter kann jedoch nicht garantiert werden für das Hochladen von beliebigen 3D Strukturen von Molekülen, die der Benutzer mittels externer, lokal installierter 3D Konstruktionsprogrammen erzeugt hat. Dies gilt insbesondere für Moleküle, die Ionen oder Metalle enthalten.

Die Eingabe der Topologie von komplexen Oligosacchariden erfolgt in von der IUPAC empfohlenen Nomenklatur der erweiterten Textnotation für Kohlenhydrate. Für lineare einfache Strukturen kann dies mittels Pull-Down-Menues erfolgen (Abbildung 22 oben), für verzweigte Strukturen und seltene Residuen durch die Eingabe der Topologie in das Textfeld. Zur Orientierung für Benutzer, die diese Notation nicht oder nur selten benutzen, ist eine direkte Verknüpfung zu einer Webseite(Knopf,SWEET') erstellt worden, in der die Topologie vieler biologisch

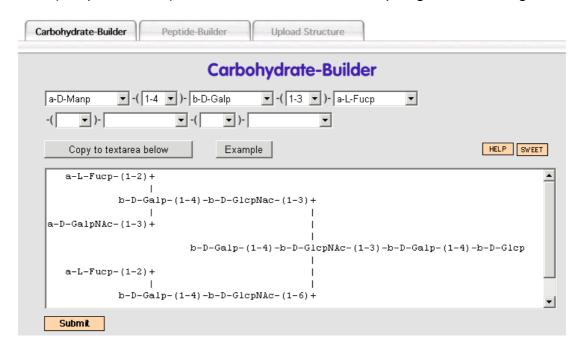


Abbildung 22) Eingabe von Kohlenhydrat Topologien mittels der erweiterten Textnotation. Einfache, lineare Strukturen werden am schnellsten mittels der Eingabe Maske (oben) eingegeben. Komplexe, verzeigte Strukturen über das Text-Eingabe Feld (unten). Um Platz zu sparen werden hier beide Eingabe-Optionen in einer Abbildung darstellt. Es besteht deshalb kein Zusammenhang zwischen den Strukturen in der oberen und unteren Eingabeebene.

wichtiger in der benötigten Schreibweise zur Verfügung gestellt wird. Durch eine einfache Kopieren/Einfügen Operation kann diese Darstellung direkt in das Textfeld geladen werden. Hat man die gewünschte Sequenz eingegeben und den "Submit' Knopf gedrückt, so erfolgt eine sehr schnelle, interaktive Konstruktion der räumlichen Struktur mittels des Programms SWEET-II, das ebenfalls in der Zentralen Spektroskopie des DKFZ entwickelt wurde.

Im 'Save new structure' Fenster wird die generierte 3D-Struktur mittels des Chime Plug-ins (links) dargestellt und kann anschließend nach Überprüfung mit Namen, Klassifizierung und Kommentar versehen in die Struktur-Datenbank des jeweiligen Benutzers abgespeichert werden. Ist dies erfolgt so kann eine neue Simulation mit dem konstruierten Molekül erfolgen.

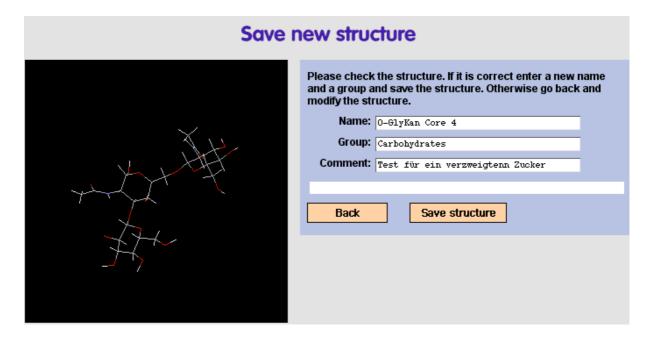


Abbildung 23) Bezeichnung und Speicherung von neu konstruierten räumlichen Strukturen in der Struktur-Datenbank, die alle Moleküle enthält, die der jeweilige Benutzer bearbeitet hat.

2.2.2.3.2 Start von Simulationen

Das "Simulation Setup' Fenster des "Experten'-Modus ähnelt weitgehend dem entsprechenden Fenster des "Beginner'-Modus mit dem wesentlichen Unterschied, dass die Werte der einzelnen Parameter der Simulation frei eingestellt werden können. Neben der Dielektrizitätskonstanten kann insbesondere die Länge einer Simulation und die Zahl der zu erzeugenden Momentaufnahmen durch den Benutzer frei gewählt werden. So kann jeder Benutzer Simulationen starten, die – je nach Größe des Moleküls und der eingegebnen Simulationsdauer – Tage bis Wochen an Rechenzeit erfordern. Dies war eines der wesentlichen Gründe dafür, dass der "Experten'-Modus nicht allgemein verfügbar gemacht wurde, da ansonsten sehr leicht die Situation entstehen könnte, dass uns wenige Benutzer, die uns zudem noch unbekannt sind, alle verfügbaren Knoten des LINUX-Clusters mit sehr langen Simulationen blockieren.

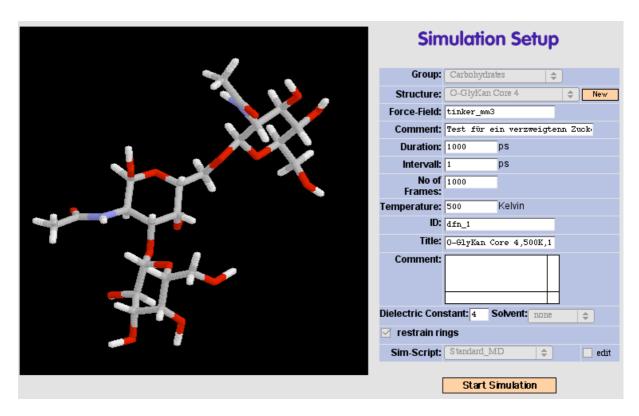


Abbildung 24) Eingabe der Simulationsbedingungen im "Experten'-Modus. Die Wiedergabe der räumlichen Struktur erfolgt in Chime-Fenster (links)

2.2.2.3.3 Bearbeitungszustand von Simulationen

Nach Aktivierung des "Start Simulation" Knopfes im "Simulation Setup" Fenster (Abbildung 24) werden alle notwendigen Daten – räumliche Struktur des Moleküls, Simulationsbedingungen und Unix-Skripte zum eigentlichen Starten der Prozesse – an den LINUX-Cluster transferiert und dort in die Warteschlange gestellt. Der aktuelle Zustand der Bearbeitung einer Berechnung kann über das "Simulation Status" Fenster abgefragt werden. Beim ersten Aufruf dieses Fensters werden dem Benutzer auch die grob abgeschätzte tatsächliche Dauer einer Simulation mitgeteilt, die auf der Grundlage von Erfahrungswerten berechnet wird.

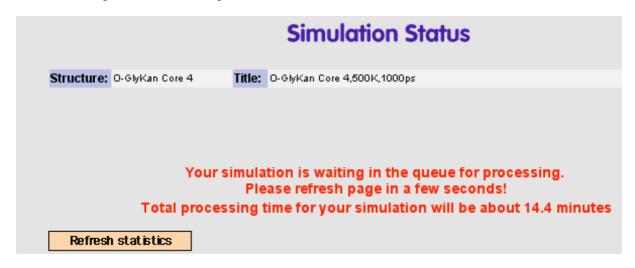


Abbildung 25) Kontrolle des Bearbeitungszustands einer Simulation unmittelbar nach dem Starten einer Simulation.

Nach dem Starten der Simulation kann der Nutzer den aktuellen Zustand einer Bearbeitung jederzeit abrufen (siehe Abbildung 26). Dies gilt ebenfalls für den zeitlichen Verlauf und die Statistik von wesentlichen Kenngrößen zur Bewertung einer Simulation wie Energie und Temperatur (siehe Abbildung 27 links) sowie für die letzte, gerade berechnete räumliche Struktur des Moleküls (siehe Abbildung 27 rechts).

Simulation Status							
Structure: 0-GlyKan Core 4	Title: 0-Gh	/Kan Core 4,500K,1	1000ps				
Job Parameters							
Temperature [K]	500		Total time [[ps]	1000		
No of frames	1000		Current tim	ie [ps]	202.6		
History frequency [ps]	1		Finished [%	1	20.3	20.3	
Dielectric constant	4		Solvent	Solvent		none	
Restraints	rings		Force Field		tinker_mm3		
Minimization		Steps		Energy		RMS	
	461		38.424		0.091		
Dynamics	Mean	Std.Dev.	Min	Max	Slope	Slope(Abs)	
Total Energy [kcal/mole]	267.78	3.68	175.37	281.23	-2.51	2.51	
Pot. Energy [kcal/mole]	154.56	3.35	106.97	169.65	-1.48	1.48	
Kin. Energy [kcal/mole]	113.22	2.07	68.4	118.84	-1.02	1.02	
Temperature [K]	499.52	9.16	301.92 524.57		-4.52	4.52	
Refresh statistics	S	how last frame	Energy Pbt		Temperature Pbt		
	Display simulation Energy Histogram			Temperature Histogram			

Abbildung 26) Information über den Bearbeitungszustands einer Simulation während der Ausführung einer Simulation. Dargestellt sind die Kennwerte der ersten 202 ps einer 1000 ps Simulation.

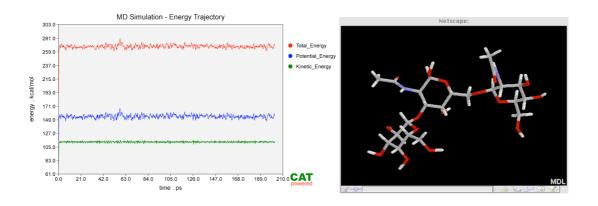


Abbildung 27) Verlauf der Energie während der ersten 202 ps der Simulation (links). Räumliche Struktur der letzten abgespeicherten Momentaufnahmen (rechts).

Nach Beendigung der Simulation hat der Benutzer die Möglichkeit, sowohl das Protokoll der eigentlichen Berechnung zu kontrollieren, das von dem jeweiligen MD Programm (TINKER in diesem Fall) erzeugt wurde, als auch das Protokoll der statistischen Auswertungen, das von dem Programm CAT erzeugt wird. In diesen Dateien wird auch festgehalten, welche Probleme bei der Ausführung einer Simulation aufgetreten sind.

2.2.2.3.4 Abspielen und Analyse von Simulationen

Das Abspielen von Simulationen sowie die Analyse der allgemeinen Charakteristika von Simulationen (Energie Temperatur) erfolgt analog dem "Beginner"-Modus über das bereits beschriebenen "Display Simulation" Fenster (Absatz 2.2.1.4. Wiedergabe von Simulationen). Auch die detaillierte Analyse von bestimmten strukturellen Merkmalen arbeitet nach dem gleichen Prinzip wie im "Beginner'-Modus und wird hier im Detail für die Untersuchung eines Trisaccharides beschrieben. Für Kohlenhydrate gibt es eine alle wesentlichen geometrischen Kenngrößen umfassende Liste von vordefinierten "Wave Definitions", die durch Aktivierung des "Get predefined" Knopfes (siehe Abbildung 12) in das editierbare Skript-Fenster geladen werden können. Die Ausführung dieser Anweisungen erzeugt eine lange Liste von Objekten, die in der Results List' wiedergegeben werden (siehe Abbildung 28). Zur besseren Handhabung der Liste können nur bestimmte Objekte aufgelistet werden, die über eine Suchfunktion ausgewählt werden können. In Abbildung 29 wurden alle Objekte vom Typ "Linkage" ausgewählt. Der Objekttyp "Linkage" ist für die Untersuchung von Kohlenhydraten besonders interessant, da er die räumliche Anordnung von benachbarten Monosaccharid-Ringen in Form von Konformationskarten beschreibt. die eine weite Verbreitung in der Literatur gefunden haben. Durch Indizierung des interessierenden Objekts und Aktivierung des entsprechenden Knopfes werden die jeweiligen Informationen angezeigt. In Abbildung 30 ist dies z.B. die Konformationskarte der 1-3 Verknüpfung zwischen zwei Zucker-Residuen.

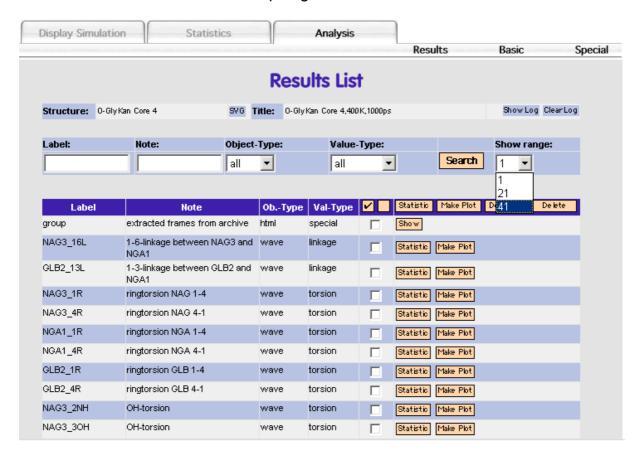


Abbildung 28) Liste der von CAT erstellten Objekte, die mittels der "Get predefined' Funktion für Kohlenhydrate definiert wurden. Die Liste enthält insgesamt 44 Objekte, von denen hier nur die ersten 10 aufgelistet sind.

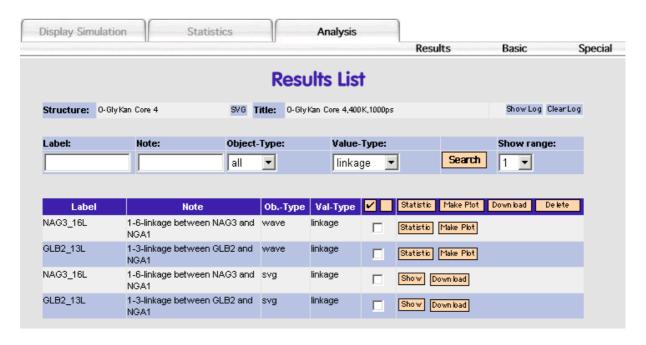


Abbildung 29) Auswahl aller Objekte von Typ ,Linkage'.

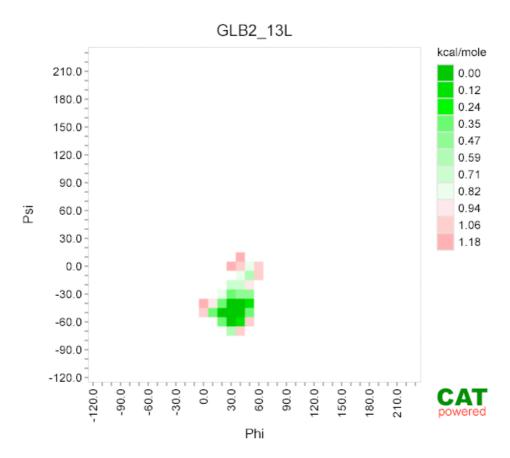


Abbildung 30) Konformationskarte (Objekte von Typ ,Linkage') der []-D-GlcpNAc-(1-3)-[]-D-GalpNAc Verknüpfung des O-Glycans Core 4. Gegeneinader aufgetragen sind die beiden Torsionswinkel Phi und Psi der glykosidischen Verknüpfung. Nur die eingefärbten Bereiche sind während der durchgeführten Simulation auch tatsächlich populiert.

2.2.2.3.5 Interface zur Definition von zu analysierenden geometrischen Kenngrößen.

Der Benutzer hat die Möglichkeit, jede beliebige geometrischen Kenngröße frei zu definieren und den zeitlichen Verlauf mit dem Programm CAT analysieren zu lassen. Zu diesem Zweck dient das Fenster 'Analysis Setup'. Zur Definition der zu analysierenden internen Koordinaten (Abstände, Winkel, Torsionwinkel, Verknüpfungen) müssen CAT die Nummern (IDs) der beteiligten Atome mitgeteilt werden. Die Übergabe erfolgt in der Form, dass der Benutzer auf das jeweilige Atom in der Moleküldarstellung (Abbildung 21 links) klickt. Die entsprechende Nummer des Atoms wird dann direkt in das frei editierbare Feld 'def_atoms' übertragen. Für diesen Datentransport zwischen der graphischen Darstellung wird die 'Live-Connect'-Funktion des Chime Plug-ins benötigt, die leider nur unter Netscape 4.7x und unter Version 7 verfügbar ist. Unter Internet Explorer erscheint beim Klicken auf das jeweilige Atom dessen Nummer in der Fußleiste des Browsers und muss dann von Hand in das 'def_atoms' Feld übertragen werden.

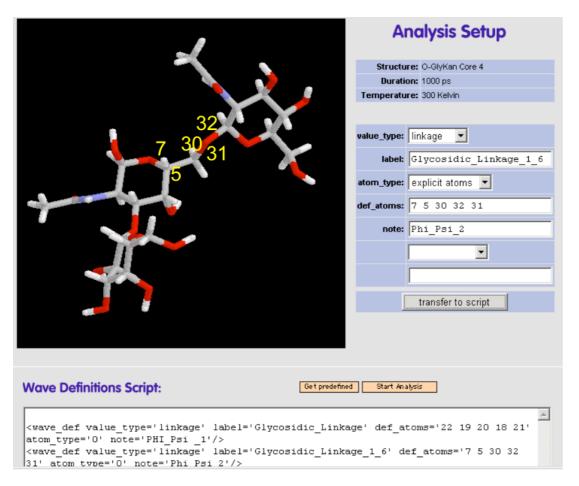


Abbildung 31) Benutzerdefinierte Eingabe von auszuwertenden inneren Koordinaten. Die Indizierung der Atome über die Eingabe ihrer Atomnummer, die erhalten wird durch Klicken mit der Maus auf das jeweilige Atom in der Moleküldarstellung. Dargestellt ist die Indizierung der glykosidischen Bindung einer 1-6 Verknüpfung.

2.2.2.3.6 Spezielle Formen der Analyse von Simulationen

In den vorhergehenden Kapiteln wurden die Möglichkeiten der Analyse von einzelnen inneren Koordinaten vorgestellt. Oft interessieren jedoch auch komplexere Fragestellungen wie etwa die Liste aller während einer Simulation auftretenden Wasserstoffbrücken-Bindungen. Die durchzuführenden Analysen hängen natürlich stark von den jeweiligen Fragestellungen ab. In der Analysesoftware CAT sind bereits diverse Optionen für komplexe Auswertungen implementiert, die allerdings bisher zum größten Teil direkt angesprochen werden müssen. Um zu demonstrieren, dass auch derartige Aufgaben mittels der Analyseoptionen von 'Dynamische Moleküle' bearbeitet werden können, wurden exemplarisch drei oft nachgefragte Auswertungen implementiert, die im Menue 'Analysis/Special' aktiviert werden können. Die "H-Bond' Analyse liefert einen kompletten Überblick über alle Wasserstoffbrücken-Bindungen, die mit einer signifikanten Häufigkeit während der Simulation auftreten. Die Option ,H-H distances erzeugt eine r-6-gewichtete Matrix von Wasserstoff-Wasserstoff-Abständen, wie sie für die Interpretation von NMR-Spektren hilfreich ist. Während die beiden genannten Optionen anwendbar für alle Klassen von Molekülen sind, wurde die Analyse der Abstände der Kohlenstoff-Alpha-Atome speziell für Proteine und Peptide implementiert.



Abbildung 32) Abstandsmatrix aller H-Atome des analysierten Trisaccharides. Auf der x- und y-Achse sind die jeweils die Bezeichnungen der vorhandenen H-Atome aufgetragen. Die Farbkodierung kennzeichnet den Mittelwert der jeweiligen H-Atome. Bewegt man die Maus über die einzelnen Elemente der Matrix, so werden in der Infozeile des Webbrowsers die jeweiligen Atome und der Abstand angezeigt.

Abbildung 32 zeigt die Abstandsmatrix aller H-Atome des analysierten Trisaccharides. Die Farbkodierung kennzeichnet den Mittelwert der Abstände der jeweiligen H-Atome während der Simulation. Von besonderem Interesse sind kurze Abstände zwischen H-Atomen, die nicht zu einem Residuum gehören (interresiduale Abstände). Eine besondere NMR-Messtechnik erlaubt es, Abstände von H-Atomen, die sich im Raum sehr nahe kommen zu bestimmen. Je mehr interresiduale H-Atome bestimmbar sind, desto genauer kann auch die räumliche Struktur eines Saccharides bestimmt werden. Um für die Interpretation der Abstände eine schnelle Zuordnung zu erhalten, wurde eine interaktive Anzeige der Abstände implementiert. Bewegt man die Maus über die einzelnen Elemente der Matrix, so werden in der Infozeile des Webbrowsers (unterer Fensterrand) die jeweiligen Atome und der Abstand angezeigt.

2.3 Allgemeines zum Portals "Dynamische Moleküle" (About)



Abbildung 33) Das , About Fenster mit allen Untermenus.

Die Hauptrubrik 'About' stellt allgemeine Daten und Hintergründe des Portals ,Dynamische Moleküle' zur Verfügung. Das Submenue ,Info' liefert eine allgemeine Beschreibung der Ziele sowie der Konzepte des Projektes. Unter ,News' (siehe Anhang) erhält der Benutzer einen schnellen Überblick über die neuesten Mitteillungen, Veränderungen und Ereignisse des Portals. ,People' zeigt die Personen, die an der Entwicklung des Portals beteiligt waren und ermöglich die jeweiligen Personen direkt per Email zu kontaktieren. Im Bereich ,Technical' wird der interne Datenfluss anhand von schematischen Darstellungen gezeigt. Eine detaillierte Darstellung der technischen Details des Projektes erfolgt im Kapitel 3 dieses Berichtes. Unter ,Gallery' befinden sich Filme, die aus verschiedenen Bereichen des Portals zusammengetragen wurden.

2.4 Informationen zum Portals ,Dynamische Moleküle' (Information)



Abbildung 34) Die Rubrik 'Information' mit allen Untermenus.

Die Rubrik ,*Information*' dient dazu, den allgemeinen wissenschaftlichen Hintergrund der Molekulardynamik darzustellen sowie Informationen anzubieten, die typische Anwendungen darstellen und das praktische Arbeiten mit den verschiedenen Funktionen des Portal erklären.

Es wurde jeweils ein Manual für den "Beginner"- und "Experten"-Modus erstellt, in denen anhand praktischer Beispiele beschrieben wird, wie die einzelnen Funktionen des Portals aktiviert werden können. Die Manuals sind im Anhang dokumentiert.

Es wurden bisher insgesamt 8 Workshops entwickelt. Darin wird dem Benutzer zum einen eine allgemeine Einführung in die Beweglichkeit und Dynamik von Molekülen vermittelt. Andererseits wird dem Benutzer anhand konkreter Moleküle und Fragestellungen dargestellt, für welche Zwecke man die verschiedenen implementierten Simulationstechniken der Moleküldynamik und die Werkzeuge zur Auswertung optimal nutzen kann.

Bisher wurden Workshops zu folgenden Themen erarbeitet:

Workshop 1: Bewegungen von Molekülen (allgemeines)

Workshop 2: Die kleine Welt der Moleküle

Workshop 3: Kleine und mittlere Ringe

Workshop 4: Konformationsanalyse des Ethans

Workshop 5: Konformationsanalyse des Butanols

Workshop 6: Konformationsanalye des n-Butans

Workshop 7: Cyclohexan-Konformationen

Workshop 8: Zeitskala molekularer Bewegungen

Dabei werden sowohl Filme von Simulationen in verschiedenen digitalen Formaten (AVI, Quicktime oder Multigif) angeboten, die zur Illustration des jeweiligen im Text gesagten dienen, als auch eher schematische Darstellungen realisiert als Flash Animationen, die den Ablauf von bestimmten Ereignissen darstellen.

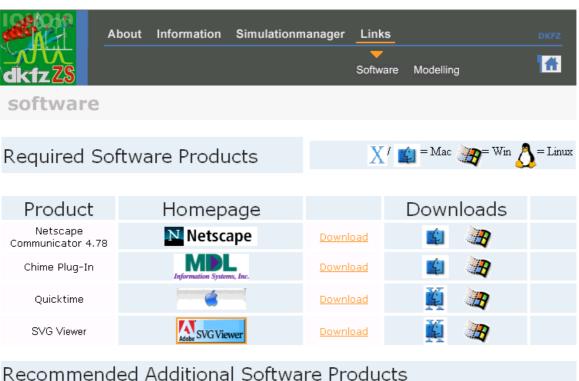
Ein Ausdruck der Texte der bisher erstellen Workshops ist im Anhang dokumentiert. Es sei hier jedoch darauf hingewiesen, dass diese Workshops auch insbesondere von Ihren Animationen und der Interaktion mit dem Benutzer leben. Dieser Aspekt kann natürlich nur sehr rudimentär durch eine gedruckte Version wiedergegeben werden.

Im Zuge der weiteren Entwicklung des Portals 'Dynamische Moleküle' sind dem Bedarf der Benutzer entsprechend in Zukunft weitere Workshops zu speziellen Themen anzubieten.

2.5 Depot der benötigten Software und verwandte Anwendungen (Link-Liste)

Das Untermenue 'Software' der Link-Rubrik bietet einen schnellen Zugang zum Download aller Softwareprodukte, die zur Erreichung der vollen Funktionalität des Portals 'Dynamische Moleküle' benötigt werden. (Netscape Browser, Chime Plug-in zur Visualisierung von Molekülen, Quicktime zum Abspielen von Videos und dem SVG Plug-in, das benötigt wird für die Darstellung der wissenschaftlichen Graphiken). Es wurde insbesondere Wert darauf gelegt, dass alle benötigten Werkzeuge frei verfügbar sind und direkt vom Netz ohne zusätzliche Kosten heruntergeladen werden können. Leider hat sich die Tendenz, dass bestimmte

Browser nur bestimmte Funktionen und Plug-ins unterstützen während des Projektes als ein größeres praktisches Problem herausgestellt. Aus diesem Grund haben wir uns entschlossen, für dieses Pilotprojekt zunächst nur eine Konfiguration zu unterstützen, in der die gesamte Funktionalität unterstützt wird. Unsere Wahl fiel dabei auf die Version 4.7. von Netscape insbesondere deshalb, weil in dieser Version die 'life connect' Funktionalität unterstützt wird, die benötigt wird, um den Transport von Daten zwischen dem Chime Plug-in und HTML-Seiten zu realisieren. Diese Funktionalität wird z.B. benutzt, wenn die Atomnummer von angeklickten Atomen automatisch in das Eingabefeld des 'Analysis Setup' Fenster transportiert werden. Zwar unterstützt auch die Version 7 von Netscape die 'life connect' Funktionalität, nur leider erkennt das aktuelle Installationsmenue von Chime nicht Netscape Version 7. Kopiert man manuell die entsprechenden Chime Dateien in den entsprechenden Netscape-Ordner, so kann Chime als Plug-in in Kombination mit Netscape 7 und Mozilla benutzt werden.



Recommended Additional Software Froducts

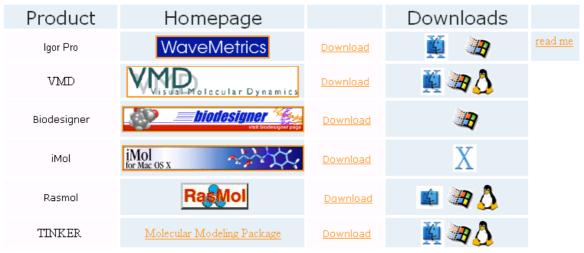


Abbildung 35) Die zum Herunterladen bereitgestellte Software für das Portal 'Dynamische Moleküle'

Darüber hinaus wird der Zugang zu einer Reihe von weiteren interessanten Programmen zur Visualisierung von Molekülen und zur weitergehenden statistischen Auswertung von Simulationen angeboten. Die Programme iMol (Mac)/Biodesigner (Windows) können zur Erzeugung Filmen genutzt werden. Die Erstellung von Filmen mit "Biodesigner" wird in Kapitel 2.2.1.1.4 beschrieben.

Auch das Programm VMD bietet die Möglichkeit, Trajektorien abzuspielen und eine Sequenz von Abbildungen zu erzeugen. Die so erstellten Bildsequenzen müssen dann mit weiteren Programmen (z.B. Qucktime Pro) zu einer Filmsequenz zusammengefügt werden. Rasmol ist ein bereits lange existierendes und sehr weit verbreitetes Programm zur Visualisierung von molekularen Strukturen. Das Plug-in Chime basiert auf Rasmol. "Igor" ist das einzige rein kommerzielle Programm in der Liste. Die Demoversion erlaubt jedoch das Laden, Visualisieren und Analysieren der Daten. Die Rohdaten von Trajektorien lassen sich über den Simulationsmanger herunterladen und mit "Igor" weiter bearbeiten.

3 Technische Realisation : interne und externe Kommunikation

3.1 Grundlegende Konzepte : Auswahl von Soft- und Hardwarekomponenten

Wesentliches Ziel des Projektes ist es, MD Simulationen als eine wichtige Computermethode zur Untersuchung der Beweglichkeit von biologischen Molekülen über ein Web-Interface allgemein zugänglich zu machen. Aus diesem Grunde wurde bei der Auswahl der benutzten Software besonderer Wert darauf gelegt, dass alle seitens des Benutzers benötigten Werkzeuge (Web-Browser und Plug-ins) möglichst kostenlos und frei zugänglich sind. Auch sollte bereits verfügbare Software z.B. für die Visualisierung der Moleküle Verwendung finden, da so aufwendige und kostenintensive Eigenentwicklungen umgangen werden können. Dabei wurde bewusst in Kauf genommen, dass im Rahmen dieses Pilotprojekts nur eine bestimmte Kombination von Werkzeugen die gesamte Funktionalität des Portals unterstützt.

Auch bei der Realisierung des Projektes auf der Serverseite wurde konsequent darauf geachtet, dass nur frei verfügbare Software (wenn möglich ,Open Source') verwendet wurde, da nur so die freie Portierbarkeit des Systems auf andere Server ohne zusätzliche Lizenzgebühren garantiert werden kann. Die Wahl fiel dabei auf ein unixbasiertes Betriebssystem mit Apache als Webserver, MySQL als relationale Datenbank und PHP als eine serverseitig interpretierte, in HTML eingebettete Skriptsprache zur dynamischen Erzeugung von HTML-Seiten und einfachen Abbildungen. Alle neu entwickelten Moleküldynamik Programme (TINKER, NAMD, GROMACS) sind für akademische Institutionen frei nutzbar. Einzig die Software zur automatischen Analyse von Simulationen – Conformational Analysis Tools (CAT) – wurde im Rahmen dieses Projektes in für das Projekt wesentlichen Teilen neu konzipiert (z.B. Betrieb als Hintergrundprozess) und kontinuierlich and an die speziellen Notwendigkeiten der Anwendung angepasst. Es gab in der .Open Source' Gemeinschaft keine Software, die die benötigte Funktionalität auch nur in Ansätzen vorweisen konnte. Der Rechenserver zur Durchführung der eigentlichen Simulationen sollte auch aus Standard Hardware-Komponenten bestehen und so ein günstiges Preis zu Leistungsverhältnis garantieren. Aus diesem Grunde viel die Wahl auf einen Cluster bestehend aus marktüblichen PCs, die unter dem Betriebsystem

LINUX laufen. Die Kommunikation der Knoten untereinander erfolgt über zwei Netzwerke mit jeweils 100 MB Ethernet Anschlüssen.

3.2 Grundlegende Konzepte : Datenfluss und Benutzerinteraktionen

Üblicherweise erfolgt bei internet-basierten Anwendungen die Ausführung der Berechnungen auf der Serverseite so schnell, dass der Benutzer nur wenige Sekunden bis etwa maximal eine Minute auf das Ergebnis zu warten hat. Wie in Abbildung 36 dargestellt ergibt sich so eine Art Kreisprozess für die klassische Web-Anwendung, bei der der Benutzer seine Anforderung/Anfragen abschickt, die eigentlich Berechnung auf dem Server erfolgt, danach eine automatische Analyse der Daten durchgeführt wird und die dann während derselben Sitzung an den Benutzer zurückgegeben wird.

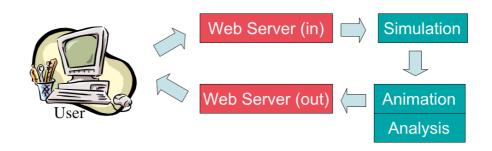


Abbildung 36) Klassische interaktive Web-Anwendung. Die Ausführung der Berechnung auf der Serverseite erfolgt so schnell, dass der Benutzer innerhalb weniger Sekunden eine Antwort erhält.

Dieser Kreisprozess der klassischen Web-Anwendung ist nicht mehr sinnvoll und praktikabel, wenn die Ausführung der eigentlichen Simulationen viele Minuten, Stunden oder sogar Tage dauern kann. Auf der anderen Seite sind die Auswertungen der Simulationen zumeist so schnell, dass sie durchaus im Rahmen einer klassischen Web-Anwendung bearbeitet werden können. Auch zeigt die Praxis, dass sich im Verlauf einer Analyse oft neue Fragstellungen ergeben, so dass ein schneller, interaktiver Zugriff auf die Simulationsdaten von großem Nutzen ist. Die in den Simulationsdaten enthaltenen Informationen sind so vielfältiger Art, dass sie zumeist nicht alle routinemäßig automatisch ausgewertet werden können. Auch ist die Art der nachgefragten Informationen abhängig von der jeweiligen wissenschaftlichen Fragestellung. Ein sehr flexibles System zu Analyse der erzeugten Primärdaten ist deshalb notwendig.

Aus den genannten Gründen wurde im Rahmen des Projektes das Konzept eines Simulationsmanagers entworfen und realisiert. Wie in Abbildung 37 dargestellt, hat der Benutzer jederzeit Zugriff auf alle Daten, die bereits erzeugt worden sind. Diese Anforderung bedeutet, dass ein System von gekoppelten Datenbank-Tabellen erstellt werden musste, das in der Lage ist, alle anfallenden Simulationen, Auswertungen und Abbildungen zu verwalten auf die der Benutzer jederzeit einfach zugreifen kann.



Abbildung 37) Der jederzeitige Zugriff des Benutzers auf alle erzeugten Daten muss realisiert werden.

In dem ursprünglichen Antrag war vorgesehen, dass der Experten-Modus auf die Untersuchung von Kohlenhydraten beschränkt wird und deshalb die Verwaltung der Daten von der bereits existierenden SWEET-DB übernommen werden soll. In Hinblick jedoch auf den beabsichtigten Ausbau des Portals auf weitere Molekülklassen, wurde das Konzept verfolgt, eine allgemeine, unabhängige Datenbankstruktur zu schaffen, die in der Lage ist, alle möglichen Klassen von chemischen Strukturen zu verarbeiten. Diese sehr aufwendigen Arbeiten gehen weit über die im Antrag zugesagten Leistungen hinaus sind jedoch sinnvoll im Hinblick auf die weitere Entwicklung des Portals.

3.2.1 Simulationsmanager

Der Simulationsmanager besteht aus einem System von gekoppelten Datenbank-Tabellen und organisiert die Verwaltung aller untersuchten chemischen Strukturen und der durchgeführten Berechnungen sowie deren Auswertungen. Die durch den Simulationsmanager unterstützte Funktionalität wurde bereits in Abbildung 3 im Kapitel 2.2. aufgelistet. Eine gesonderte Verwaltung der Benutzer ist notwendig, da nur so garantiert werden kann, dass jeder Benutzer nur in der Lage ist, seine eigenen Daten einsehen. Dies ist eine wesentliche Bedingung für den Einsatz und die Akzeptanz des Portals durch Wissenschaftler, die konkrete Forschungsprojekte durchführen wollen. Die Rohdaten der Simulationen, die je nach Molekül und Simulationsdauer bis zu einigen GigaByte groß werden können, werden als einzelne Dateien in Ordnern des Servers abgelegt und durch die entsprechenden Verweise in den Tabellen des Simulationsmanagers referiert. Die Handhabung der verschiedenen Komponenten des Simulationsmanagers durch den Benutzer wurde bereits ausführlich in Kapitel 2 beschrieben.

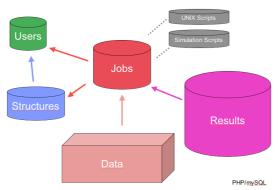


Abbildung 38) Komponenten des Simulationsmanagers.

3.2.2 Interner Datenfluss

Entsprechend den Benutzer-Anforderungen wird über den Simulationsmanager entweder eine neue Simulation auf dem LINUX-Cluster gestartet oder neue Auswertungen von bereits durchgeführten Simulationen durchgeführt. Die gesamte Interaktion mit den Benutzern, die Verwaltung der Daten und die Erzeugung von graphischen Darstellungen erfolgt auf dem Zentralen Server (grün in Abbildung 39) des Projektes erfolgt. Hier werden auch alle Dateien erstellt, die alle benötigten Angaben im richtigen Format enthalten wie sie benötigtet werden, um eine Molekulardynamik-Simulation zu starten. Die eigentlichen Simulationen, die bei weitem den rechenintensivsten Teil der Anwendung darstellen, werden auf dem LINUX-Cluster ausgeführt, der aktuell mit 18 Prozessoren bestückt ist. Die Rechnungen können je nach Größe des Moleküls und der gewählten Simulationsdauer bis zu mehreren Tagen dauern und Dateien im Bereich von einigen GigaByte produzieren. Die Auswertung der allgemeinen statistischen Kenngrößen. die für jede Simulation ausgewertet werden, erfolgt während und nach Beendigung der Berechnung auf dem LINUX-Cluster durch die Conformational Analysis Tools. Die Ergebnisse der Auswertungen werden in der Results-Tabelle des Simulationsmanagers abgelegt. Nach Beendigung der Simulation werden die Rohdaten auf den zentralen File-Server kopiert. Die Speicherplatten des zentralen Servers sind auch dem LINUX-Cluster zugänglich, so dass auch bei weiteren Auswertungen auf die Simulationsrohdaten zugegriffen werden kann. Die Erzeugung von graphischen Darstellungen aus Trajektoriendaten (waves) erfolgt mittels der CAT-Software, allerdings wird hierzu die auf dem Server installierte CAT-Version benutzt.

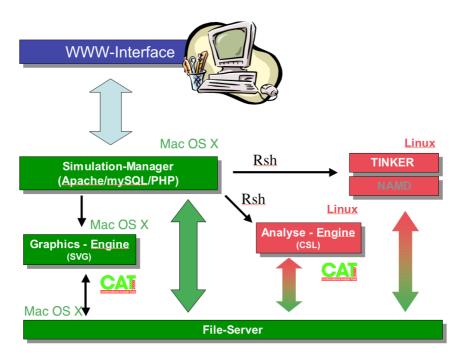


Abbildung 39) Interner Datenfluss des Portals 'Dynamische Moleküle'. Die Farbkodierung zeigt die jeweils benutzte Hardware-Plattform an: Blau: Browser des Benutzers; grün: Zentraler Server (Mac OS X); rot. LINUX-Cluster

Angesichts der schon weiter oben diskutierten Anforderungen an Rechenzeiten und Hardware-Ressourcen ist es im Rahmen dieses Pilotprojektes nicht sinnvoll, Simulationen von großen Proteinen in einer wässrigen Umgebung über das Web-

Interface zu starten. Diese Simulation würden die verfügbaren Rechner wahrscheinlich für Tage, wenn nicht Wochen blockieren. In der gegenwärtigen Konfiguration ist eine Nutzung des Portals "Dynamische Moleküle" für kleinere Moleküle bis zu einer maximalen Größe von etwa 1000 Atomen und ohne wässrige Umgebung sinnvoll.

3.2.3 Erzeugung aller benötigten Daten zum automatischen Starten von Moleküldynamik Simulationen.

Die frei verfügbare Software zur Durchführung von Moleküldynamik Simulationen bietet meist keine komfortable grafische Unterstützung an für Generierung von Molekülen, die Auswahl von Simulationsskripten, das Starten und die Kontrolle von Simulationen sowie für die Auswertung von der produzierten Daten. Der Benutzer muss sich in der Regel erst durch viele Seiten einer Benutzeranleitung hindurcharbeiten, bevor er in der Lage ist, erstmalig eine sinnvolle Simulation starten zu können. Das Portal Dynamische Moleküle zielt darauf ab, möglichst viele der auf diesem Weg notwendigen Schritte zu automatisieren.

Benötigt wird für jede Simulation die Startstruktur des zu untersuchenden Moleküls. Drei unterschiedliche Wege der Struktureingabe sind realisiert:

- a) Abrufen von kompletten Strukturen aus einer Datenbank. Diese Option ist realisiert im "Beginner'-Modus.
- b) Erzeugen von Strukturen mittels eines textbasierten Molekül-Editors. Realisiert für die Konstruktion von komplexen Oligosacchariden und Peptiden im "Experten"-Modus

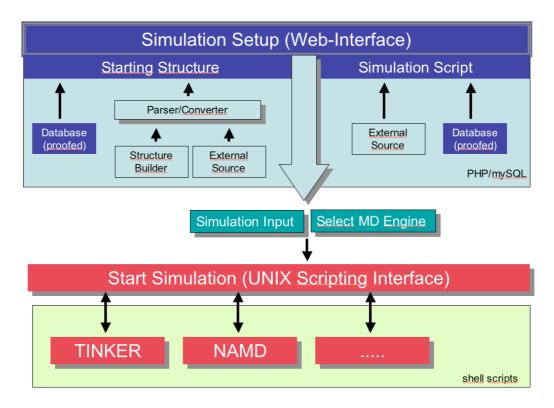


Abbildung 40) Benötigte Daten und Datenfluss zum Starten einer Molekular-Dynamik Simulation

c) Einlesen von beliebigen Molekülen, die mit externen Molekül-Editoren generiert wurden.

Die reinen Strukturdaten der Moleküle, die zumeist in Form von so genannten Verknüpfungstabellen vorliegen – Element, Atom-Koordinaten, Verknüpfungen – müssen dem Kraftfeld zugeordnet werden, mit dem die jeweilige Berechnung erfolgen soll. Leider verwendet jedes Kraftfeld seine ihm eigenen Definitionen und Bezeichnungen, so dass zunächst für jedes einzelne Atom zunächst eine Analyse der atomaren Nachbarschaft zu erfolgen hat. Aufgrund dieser Analyse kann dann der Atom-Typ bestimmt werden und eine Zuordnung zu den Kraftfeldparametern des jeweiligen Kraftfeldes erfolgen. Die Vielfältigkeit sowohl der zu berücksichtigenden chemischen Umgebungen als auch die Zuordnungen zu den jeweiligen Kraftfeldparametern, die von den unterschiedlichen Kraftfeldern gefordert werden, hat uns dazu bewogen, dass wir uns im Rahmen dieses Pilotprojektes auf zwei Molekülklassen - Oligosaccharide und Peptide – und drei Kraftfelder (TINKER-MM3, TINKER-AMBER und NAMD-CHARMM) konzentrieren. Zu diesem Zweck wurde eigens ein Konverter entwickelt, dessen Design so flexibel gehalten ist, dass er mit vertretbarem Aufwand auch für weitere Kraftfelder angepasst werden kann.

Die zweite benötigte Information sind die physikalischen Bedingungen wie Temperatur, Druck etc., unter denen eine Simulation durchgeführt werden soll. Diese können entweder direkt über eine Abfragemaske und/oder aus einer Datenbank mit vordefinierten Simulationsskripten abgerufen werden.

Beide Dateien – Molekülbeschreibung und Simulationsskripte – werden auf den LINUX-Cluster transferiert. Dort werden über einen Remote-Shell-Aufruf die entsprechenden Skripte zum Starten der Molekulardynamik-Programme – zur Zeit werden NAMD und TINKER unterstützt – ausgeführt.

3.2.4 Conformational Analysis Tools (CAT)

Eine zentrale Aufgabe des Projektes war die Entwicklung von Konzepten der interaktiven internet-basierten Analyse von Moleküldynamik Simulationen. In der Literatur sind bisher keine derartigen Ansätze beschrieben und auch intensives Studium der verfügbaren "Open Source" Angebote ergab keine geeignete Software, die die benötigten Funktionalitäten vorweisen konnte. Glücklicherweise konnten wir auf die von Dr. Martin Frank im Rahmen seiner Dissertation entwickelte Software "Conformational Analysis Tools" (CAT 2000) zurückgreifen, die in ihrem Funktionsumfang den Anforderungen des Portals "Dynamische Moleküle" weitgehend entsprach. Anpassungen waren jedoch notwendig, damit die Software effizient in Kombination mit einer Web-Anwendung eingesetzt werden konnte. Aus diesem Grunde wurde eine neue XML-basierte Skriptsprache – CAT Skript Language (CSL) genannt – implementiert, mittels derer es möglich ist, die auszuwertenden inneren Koordinaten dem Programm CAT mitzuteilen. Die einzelnen in dem Skript aufgelisteten Anweisungen müssen alle Informationen enthalten, die notwendig sind um eine gewünschte Aktion durchzuführen. Soll z.B. der zeitliche Verlauf des Abstandes von zwei Atomen gemessen werden und davon gleich eine graphische Darstellung erstellt werden, so müssen neben der Definition der Atome auch z.B. die Anweisungen für die Erstellung der graphischen Darstellungen vorhanden sein. Im Anhang ist sowohl ein Tutorial als auch das Reference Manual von CAT dokumentiert.

3.3 Auswahl der Hardwarekomponenten

Zur Ausführung von drei unterschiedlichen Anforderungen und Funktionen waren Hardware-Entscheidungen zu treffen:

- a) Web-Server
- b Rechenserver zur Durchführung der Moleküldynamik Simulationen
- c) Erstellung von Videosequenzen.

Wie schon in Kapitel 2.2.1.1.4 ,*Erstellung von Filmen*' dargestellt, hat es sich im Verlauf des Projektes aus verschiedenen Gründen als sinnvoll erwiesen, dass eine serverseitige Erzeugung von Videosequenzen nicht besonders praktikabel ist und die in Kapitel 2.2.1.1.4 beschriebene Alternative viele praktische Vorteile bietet. Aus diesem Grunde wurde auf die Anschaffung eines speziellen Servers zur Erstellung von Videosequenzen verzichtet.

3.3.1 Web-Server

Weltweit werden überwiegend UNIX-basierte Systeme als Web-Server eingesetzt. Sie zeichnen sich durch eine hohe Stabilität, kostenlose Verfügbarkeit der benötigten Softwarekomponenten, rasche Entwicklung von weiteren "Open Source" Angeboten, und einem besonders günstigen Preis-Leistungsverhältnis bei der Hardware aus. Vielfach implementiert werden so genannte L(W)AMP-Systeme (LINUX oder Windows als Betriebssystem, Apache als Web-Server, MySQL als Datenbanksystem und PHP als Hypertextpreprozessor zur dynamischen Erzeugung von HTML-Seiten). Es liegt nahe, dass ein Projekt wie "Dynamische Moleküle", welches man in die Kategorie "Open Access" Projekt einordnen kann, auch versucht, auf den entsprechenden "Open Source" Angeboten aufzusetzen. Die Festlegung auf "Open Source" Software wird die Portierung der Anwendung als "in house" Anwendung vereinfachen, da keine Probleme bei den Lizenzen der verwendeten Software auftreten sollten.

Die Wahl des Web-Servers fiel auf einen Rechner der Firma Apple unter dem Unix Betriebssystem *Mac OSX*, welches die volle oben beschriebene Funktionalität eines Web-Servers besitzt. Der Vorteil dieses Betriebssystems gegenüber einem traditionellen LINUX-PC ist darin zu sehen, dass ein Zugriff auf praktisch alle unter UNIX entwickelten Softwarekomponenten (Apache, MySQL, PHP etc.) erfolgen kann und gleichzeitig, auf demselben Desktop - ohne den Rechner neu starten zu müssen - viele kommerzielle Programme z.B. zur Erstellung von Webseiten, Präsentationen und zur Bearbeitung von Bildern verfügbar sind. Zumindest für die Phase der Entwicklung eines Projektes haben sich diese erweiterten Möglichkeiten von Mac OSX als sehr hilfreich erwiesen. In der täglichen Praxis haben sich dadurch viele notwendige Arbeitschritte wesentlich vereinfacht. Die spätere Portierbarkeit der gesamten Anwendung auf einen LINUX-Rechner bleibt dabei jederzeit gewährleistet, da Mac OSX ein komplettes UNIX-Betriebsystem (BSD) ist. Der Mac OSX Rechner konnte problemlos in die Rechnerlandschaft der anderen Web-Server unserer Arbeitsgruppe integriert werden, die unter LINUX laufen. Der Rechner und läuft seit nunmehr fast zwei Jahren stabil und ohne größere Probleme.

3.3.2 LINUX Cluster

Aufgabe des LINUX-Clusters ist, die rechenintensiven Moleküldynamik Simulationen und deren Auswertungen mit dem Programm CAT auszuführen. Mit den Programmen NAMD, GROMACS und TINKER stehen Softwareprodukte frei zur Verfügung, deren Algorithmen schon im Hinblick auf eine Cluster-Hardwarearchitektur optimiert worden sind. Die Kommunikation der einzelnen Knoten erfolgt über Standard 100 MB Netzwerkanschlüsse.

Unter diesen Randbedingungen galt es, die optimale Hardware Konfiguration mit dem besten Preis-Leistungsverhältnis für die beabsichtige Aufgabe zu finden. Dabei sollte das komplette System schlüsselfertig geliefert und installiert werden, so dass die im Rahmen dieses Projektes notwendigen Installationen und Anpassungen auf ein Minimum reduziert werden konnten.

Von anderen ähnlich gelagerten Projekten zur Optimierung von Rechnerarchitekturen für bestimmte wissenschaftliche Projekte, haben wir die am DKFZ gängige Praxis übernommen und 'Benchmarks' der geplanten Anwendung auf der angestrebten Rechnerarchitektur durchgeführt. Die so erzielten Rechenzeiten sind ein objektives Bewertungskriterium, mit dem das Leistungsvermögen der jeweiligen Konfiguration einfach und praxisnah bewertet werden kann.

Unser Vorgehen war wie folgt: wir haben zwei typische Anwendungen (jeweils eine mit dem Programm NAMD und TINKER) zusammengestellt und die Anbieter gebeten, die Benchmarks auf der Architektur auszuführen, die sie uns zum Kauf empfehlen. Letztendlich sind dann nur zwei Anbieter von LINUX-Clustern übrig geblieben, die bereit waren die geforderten Benchmarks durchzuführen. Beide Firmen haben sich auf Cluster-Architekturen im Bereich von wissenschaftlich technischen Anwendungen spezialisiert. Es erwies sich, dass das beste Preis-Leistungsverhältnis für die angestrebte Anwendung mit Dual-Prozessoren der Firma Athlon erzielt werden konnte. Beide Lieferanten boten uns Cluster mit 9 bzw. 10 Knoten an. In einem längeren Auswahlprozess, der im Anhang dokumentiert ist, wurden dann die einzelnen Hardware-Komponenten des Clusters auf der Grundlage von Testberichten optimiert. Unsere Wahl fiel schließlich auf ein System der Firma Megware mit insgesamt 10x2 Prozessoren.

Der Cluster wurde im September 2003 "schlüsselfertig" geliefert. Alle benötigte Software war installiert, so dass wir bereits am Tage nach der Aufstellung des Rechners die ersten Simulationen starten konnten. Der Cluster wurde in einem klimatisierten Raum der Zentralen Datenverarbeitung des DKFZ aufgestellt und kann komplett über eine graphisches Management System verwaltet werden. Auch technische Kenngrößen wie etwa die Temperatur jedes einzelnen Prozessors und Auslastung können über das Netz abgefragt werden. Der Cluster musste bisher nur einmal wegen Wartungsarbeiten an der Klimaanlage heruntergefahren werden. Hardware-Probleme sind bisher keine aufgetreten.

LINUX Cluster

Technische Daten

10 Rechner: 1 Master und 9 Knoten

Jeweils 2 CPUs (AMD Athlon MP2000+)

Jeweils 1GB RAM

Jeweils 2x40 GB Festplatte

(Master 2x80 GB)

2 Netzwerke: 1 für Rechenprozesse,

1 für Kontroll-und Administrationsprozesse



Abbildung 41) Der für das Projekt 'Dynamische Moleküle' installierte LINUX-Cluster geliefert von der Firma MegWare

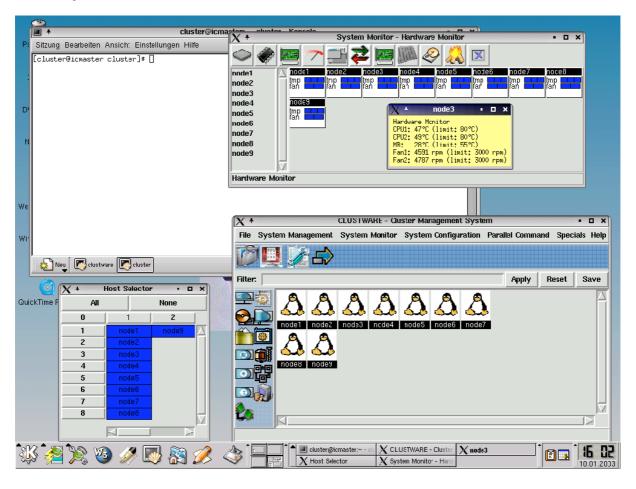


Abbildung 42) Kontrollfenster des LINIX-Clusters. Der Zustand, die Auslastung sowie die Temperatur werden für jeden einzelnen Prozessor überwacht.

4 Bekanntmachung des Portals und Benutzerresonanz

Primäre Zielgruppen der Methoden zur Untersuchung des dynamischen Verhaltens von Molekülen, die durch das Portal "Dynamische Moleküle" allgemein zugänglich gemacht wurden, waren Wissenschaftler, Studenten und Schüler. Es wurden intensive Anstrengungen unternommen, das Projekt durch Vorträge auf nationalen und internationalen Konferenzen bekannt zu machen. Eine Auflistung der gehaltenen Vorträge befindet sich im Anhang. Publikationen in wissenschaftlichen Zeitschriften und Mitteilungsorganen der Fachgesellschaften sind ein weiteres probates Mittel, auf die Funktionen einer wissenschaftlichen Anwendung hinzuweisen. Im Anhang sind die bisher erschienenen Publikationen erschienen. Weitere Publikationen sind in Vorbereitung beziehungsweise geplant. Diese beinhalten die Durchführung einer kompletten wissenschaftlichen Anwendung zu beschreiben, die von PD Dr. Hans-Christian Siebert und Prof. Hans-Joachim Gabius von der LMU München durchgeführt wurde. Untersucht wurde die Einschränkung der konformativen Vielfalt von verzweigten Zuckerstrukturen, die Träger biologischer Information sind wie z.B. den verschiedenen Blutgruppen Antigene. Es ist geplant, diese Untersuchung bei der Zeitschrift , Carbohydrate Research' einzureichen. Der Konzeptentwurf einer inhaltlichen Beschreibung des "Experten"-Modus ist ebenfalls fertig und soll in den nächsten Wochen finalisiert werden. Es ist beabsichtigt, diesen bei der Zeitschrift "Journal of Molecular Modeling" einzureichen. Darüber hinaus ist noch geplant, eine Darstellung des Projektes in einer pädagogisch orientierten Zeitschrift in deutscher Sprache zu platzieren, um so gezielt Lehrer im deutschsprachigen Raum anzusprechen.

Schüler und Studenten werden über Publikationen in Fachzeitschriften sicherlich weniger angesprochen. Um eine Rückmeldung von diesen Zielgruppen zu erhalten wurde je eine Fortbildung im DKFZ und an der Martin Luther Universität in Halle (Prof. Bögel) durchgeführt. Eine weitere Demonstration des Portals mit praktischen Übungen fand mit zwei Klassen der Oberstufe der Goetheschule in Neu-Isenburg (Hessen) statt. Von allen diesen Veranstaltungen sind Berichte und Fotos über die News-Funktion des Portals einzusehen.

Dr. Martin Frank wurde – als für die Entwicklung des Portals 'Dynamische Moleküle' federführender Wissenschaftler – im Rahmen des 17. Molecular Modelling Workshops mit dem Web-Award der Molecular Graphics und Modelling Society (Deutschsprachige Sektion) ausgezeichnet.

4.1 Publikationen

Frank, M: Bioinformatik für Jedermann, DFN Mitteilungen 61, März 2003

Frank, M. *Dynamik von biologischen Makromolekülen per Internet,* Mitteilung in GDCh: Nachrichten aus der Chemie, 3, 2003

Frank M, Gutbrod P, Hassayoun C, Von Der Lieth CW.

Dynamic molecules: molecular dynamics for everyone. An internet-based access to molecular dynamic simulations: basic concepts.

J Mol Model (Online). 2003 Oct;9(5):308-15. Epub 2003 Aug 08.

4.2 Vorträge und Poster auf wissenschaftlichen Tagungen

11thEuropean Carbohydrate Symposium - Eurocarb XI; 2.-7. September 2001; Lissabon; Portugal. *Conformational Analysis Strategies to detect Solvent Effects based on Molecular Dynamics Simulations*

Euroworkshop II on Interdisciplinary Perspectives of Diagnostic, Pathology, Cell Biology, and Morphometry; 5. - 7. Oktober 2001; Brüssel, Belgien. *Glycosciences and the internet: new developments*

- 8. Jahrestagung der luK-Initiative Information und Kommunikation der wissenschaftlichen Fachgesellschaften, 11.-13. März 2002, Ulm; *DFN-Projekt ,Dynamische Moleküle': Darstellung und Analyse der Dynamik von biologischen Makromolekülen per Internet*
- 16. Molecular-Modeling Workshop, 7.-8. Mai 2002, Darmstadt: "Dynamische Moleküle Darstellung und Analyse der Dynamik von biologischen Makromolekülen per Internet" Beginner-Modus

21st International Carbohydrate Symposium; 7.-12. Juli 2002; Cairns; Australien Poster: 'Dynamic Molecules' - The Representation And Analysis Of The Dynamics Of Carbohydrates Over The Internet (Beginner Mode)

3rd Pan Pacific Conference (3PPC) on Sialo-glycoscience and other Novel Forms of Glycosylation; 14.-17. Juli 2002, Broadbeach, Gold Coast, Australien Conformational Analysis Of N-Glycans Over The Internet: Structure, Dynamics And In Silico Glycosylation

GDCh-Workshop Computers in Chemistry, (CIC); 10.-12.November 2002, Berlin , Dynamische Moleküle' - Darstellung und Analyse der Dynamik von biologischen Makromolekülen per Internet (Beginner-Modus)

17. Molecular-Modeling Workshop, 27.-28. Mai 2003, Erlangen Vortrag+ Präsentation: "Dynamische Moleküle" - Darstellung und Analyse der Dynamik von biologischen Makromolekülen per Internet (Experten-Modus)

High-Performance Computing at the DKFZ, 23. Juni 2003, Heidelberg Running And Analyzing Molecular Dynamics Simulations of Biological Important Molecules Over The Internet

12thEuropean Carbohydrate Symposium - Eurocarb XII; 7.-11. Juli 2003, Grenoble, Frankreich Poster: 'Dynamic Molecules': The Representation And Analysis Of The Dynamics Of Carbohydrates Over The Internet (Expert Mode)

GDCh-Workshop 'Computers in Chemistry' (CIC); 7.-9. Oktober 2003, München Poster: ,Dynamische Moleküle' - Darstellung und Analyse der Dynamik von biologischen Makromolekülen per Internet

4.3 Seminare und Fortbildungen

DKFZ Fortbildung, 12.-13. Februar 2003, Heidelberg Vortrag und Workshop: *Molekular-Dynamik von biologischen Molekülen*

DFN-Verein, 11. März 2003, Berlin

Vortrag: Moleküldynamik Simulationen - was ist das?

Präsentation: "Dynamische Moleküle" - Darstellung und Analyse der Dynamik von biologischen Makromolekülen per Internet

Universität Halle, 24.-25. April 2003, Halle

Vortrag und Workshop: "Dynamische Moleküle" - Darstellung und Analyse der Dynamik von biologischen Makromolekülen per Internet

Goethe-Gymnasium, 11. Juni 2003, Neu-Isenburg

Workshop: 'Dynamische Moleküle' - Darstellung und Analyse der Dynamik von biologischen Makromolekülen per Internet

Symposium: Molecular Modelling of Carbohydrates, 6. Juli 2003, Grenoble, Frankreich. Vortrag und Präsentation: *Running And Analyzing Molecular Dynamics Simulations of Carbohydrates Over The Internet*

Symposium: Glykane und Glykomimetika in Medizin und Pharmazie, 14. November 2003, Charité Berlin –Universitätsmedizin. Poster: *Running And Analyzing Molecular Dynamics Simulations of Carbohydrates Over The Internet*

4.4 Nutzungsstatistiken

Ein wichtiges Kriterium, um die Verbreitung und Akzeptanz einer Anwendung zu dokumentieren und zu bewerten, sind Statistiken ihrer Nutzung. Diese haben wir seit März 2003 geführt. Seit diesem Zeitpunkt ist der "Experten"-Modus im Netz verfügbar. Abbildung 43 zeigt die Anzahl der Anfragen und die Menge der heruntergeladenen Daten. In der Statistik berücksichtigt werden nur Aufrufe von externen Benutzern. Nicht gezählt werden interne Aufrufe sowie Anfragen von Suchmaschinen wie etwa Google, Altavista, Yahoo etc.. Die Spitzen im April und Juli sind eindeutig auf intensive Werbung vor und während zweier Konferenzen (Modelling Workshop in Erlangen und EuroCarb in Grenoble) zurückzuführen. Abbildung 44 dokumentiert, dass die Mehrzahl der Aufrufe zwar aus Deutschland kommen, aber das Portal auch schon einen auten internationalen Bekanntheitsgrad erreicht hat. Abbildung 45 zeigt die 30 Organisationen/Internetanbieter mit den häufigsten Zugriffen. Auch hier zeigt sich, dass trotz unserer intensiven Bemühungen auf internationalen Konferenzen das Portal bekannt zu machen, die häufigsten Benutzer noch aus dem deutschsprachigen Raum kommen. Vielleicht spiegelt sich in diesen Zahlen auch wieder, dass der "Beginner"-Modus und die Workshops in Deutsch geschrieben sind. Eine zukünftige Entwicklungsarbeit wird sein, das Portal komplett zweisprachig zu gestalten.

Each unit (_) represents 50 requests for pages or part thereof.

```
month: reqs: pages:
-----: ----: Mar 2003: 2193: 333: Apr 2003: 7784: 1554: May 2003: 4644: 633: Jun 2003: 2628: 454: Jul 2003: 5377: 1858: Aug 2003: 1562: 492: Sep 2003: 4229: 638: Oct 2003: 2948: 603:
```

Busiest month: Jul 2003 (1,858 requests for pages).

Abbildung 43) Nutzungsstatistik des Portals für die Zeit von März bis Oktober 2003

Listing domains, sorted by the number of requests.

```
reqs: %bytes: domain
----: -----: -----
13947: 16.53%: .net (Network)
8287: 48.27%: .de (Germany)
 1943: 5.76%: .com (Commercial)
 1574: 5.58%: .fr (France)
 325: 2.80%: .nl (Netherlands)
  270: 1.02%: .at (Austria)
  262: 1.60%: .edu (USA Educational)
  262: 0.30%: .se (Sweden)
  210: 1.17%: .fi (Finland)
  209:
       0.74%: .uk (United Kingdom)
  181:
       0.43%: .cz (Czech Republic)
       0.38%: .be (Belgium)
  154:
       0.41%: .pl (Poland)
  145:
       0.52%: .ch (Switzerland)
  140:
       0.27%: .es (Spain)
  122:
       0.34%: .hu (Hungary)
  116:
       0.03%: .ru (Russia)
  93:
       0.46%: .it (Italy)
  86:
       0.14%: .no (Norway)
  84:
  78: 0.34%: .jp (Japan)
   68: 0.51%: .au (Australia)
  58: 0.32%: .il (Israel)
```

Abbildung 44) Nutzungsstatistik des Portals aufgeteilt nach Ländern für die Zeit von März bis Oktober 2003

Listing the first 30 organisations by the number of requests, sorted by the number of requests.

```
reqs: %bytes: organisation
----: -----: ------
13329: 15.81%: t-dialin.net
 1635: 2.99%: t-ipconnect.de
1460: 5.17%: aol.com
1196: 3.42%: grenet.fr
 670: 4.49%: uni-halle.de
 629: 2.66%: uni-erlangen.de
 451: 6.50%: fh-regensburg.de
 399: 0.90%: fz-borstel.de
 331: 2.16%: tu-darmstadt.de
 306: 1.73%: dfn.de
 274: 0.29%: bellsouth.net
 249: 0.43%: uni-stuttgart.de
 243: 4.52%: ruhr-uni-bochum.de
230: 1.67%: cnrs.fr
 201: 0.98%: ac.at
201: 2.73%: kun.nl
196: 2.00%: uni-heidelberg.de
 196: 0.73%: ac.uk
 193: 0.39%: t-online.de
 180: 0.43%: muni.cz
 166: 3.28%: tu-bs.de
 159: 0.10%: pppool.de
 154: 0.18%: mh-hannover.de
 154: 0.38%: ac.be
 153: 1.12%: csc.fi
 152: 0.03%: uni-karlsruhe.de
 138: 0.04%: attbi.com
 135: 0.28%: su.se
 134: 2.24%: mpq.de
4950: 21.41%: [not listed: 242 organisations]
```

Abbildung 45) Nutzungsstatistik des Portals für die 30 Organisationen/Internetanbieter mit den häufigsten Anfragen.

5 Ausblick

Mit dem Pilot-Projekt "Dynamische Moleküle" konnte gezeigt werden, dass Moleküldynamik Simulationen und deren detaillierte Auswertung komplett per Internet durchgeführt werden können. Es ist zu erwarten, dass mit dieser Entwicklung und weiteren nachfolgenden Projekten die Darstellung und Untersuchung des dynamischen Verhaltens von Molekülen einen breiteren Raum in Forschung und Lehre einnehmen wird.

In der kurzen Zeit seiner weltweiten Verfügbarkeit hat das Portal national wie international schon große Aufmerksamkeit erhalten. Die wissenschaftlichen Kontakte, die sich über die weltweite Verfügbarkeit des Portals ergaben, waren sehr ermutigend und haben uns in der Richtigkeit unseres Ziels bestärkt. wissenschaftliche Kommunikation auf hohem Niveau effektiv gestalten zu können. Das Portal ist zu einem festen Bestandteil unseres Web-Angebots geworden und wird in der gegenwärtigen Form mindestens weitere drei Jahre zur Verfügung stehen. Kürzlich haben wir aus DKFZ-Mitteln 4 weitere Prozessoren bestellt, so dass die maximale Ausbaustufe der bisherigen Installation des LINUX-Clusters erreicht ist. Eine inhaltliche Weiterentwicklung des Portals ist auf verschiedenen Ebenen denkbar. Findet der mehr pädagogisch ausgerichtete "Beginner"-Modus einen breiten Eingang in die Lehre an Schulen und Hochschulen, so werden sich sicherlich Kollegen finden, die spezifische Untersuchungen mit den Molekülen ihrer Wahl erarbeiten und dafür auch die entsprechenden Workshops erstellen. Alle inhaltlichen Vorschläge und Wünsche nach weiteren Molekülen und Optionen der Auswertung werden wir versuchen nach Möglichkeit umgehend zu erfüllen. Eine weitere Entwicklung des "Beginner'-Modus unsererseits ist momentan nicht beabsichtigt. Der Schwerpunkt unserer eigenen Entwicklungen wird in der Verfügbarmachung von weiteren Molekülklassen, komplexeren Simulationsskripten, größeren molekularen

Ensembles und weiteren Optionen zur Analyse liegen. Dies zielt insbesondere darauf ab, MitarbeiterInnen des DKFZ einen effizienten Zugriff zu diesen Methoden zu eröffnen. Klar ist, dass der dafür benötigte Anstieg an Rechenleistung mit dem momentan installierten LINUX-Cluster nicht abgedeckt werden kann. Klar ist auch, dass seitens der Zentralen Spektroskopie des DKFZ keine Absicht besteht, einen entsprechenden allgemeinen externen Service gegen Bezahlung anzubieten. Um den steigenden Bedarf an Rechenleistung abzudecken, bieten sich verschiedene Optionen an. Zum einen könnte "Dynamische Moleküle" so erweitert werden, dass es als Portal für GRID-Computing Projekte genutzt werden kann. Die wesentlichen Elemente der Benutzer-Interaktionsebene eines derartigen Portals sind bereits vorhanden. Was momentan fehlt ist die Aufteilung der einzelnen Prozesse auf die verschiedenen verteilten Rechner und das Abrechnungssystem. Eine andere Option besteht darin, das existierende Portal als Intranet-Anwendung bei verschiedenen Institutionen zu installieren, die dann die Simulationen auf ihren hauseignen Rechnern durchführen können. Der hierfür notwendige Aufwand unsererseits ist überschaubar und würde sich im Wesentlichen auf die Erstellung von Installationsroutinen beschränken. Denkbar wäre auch noch, das Portal so zu organisieren, dass es einfach als lokale Einzelplatz-Anwendung auf einem PC installiert werden kann.

Der hier vorgestellte Prototyp des Portals "Dynamische Moleküle" hat noch ein enormes wissenschaftliches Entwicklungspotential, das wir entsprechend unseren zeitlichen Möglichkeiten versuchen werden zu entwickeln. Die Entwicklungen werden sich einerseits durch die an uns von den externen Benutzern herangetragenen

Wünschen und Projekten orientieren. Anderseits werden sicherlich hausinterne Anforderungen die weitere Entwicklung entscheidend beeinflussen. "Dynamische Moleküle" bleibt also ein lebendiges System, das kontinuierlich von uns entsprechend wissenschaftlichen Anforderungen weiter entwickelt werden wird. Es ist zu hoffen, dass es seine Vorreiterrolle als erste vollständig internet-basierte Implementierung, die Moleküldynamik Simulationen über das Netz verfügbar macht, halten und ausbauen kann, wie wohl zu erwarten ist, dass es weitweit weitere konkurrierende Projekte dieser Art geben wird.

6 Verwendete Technologien

6.1 Apache

Ein Freeware Web Server, der bei über der Hälfte aller Websites eingesetzt wird und auf dem NCSA -Web-Server basiert. Mit viel Funktionalität, einfachem modularen Aufbau und einer Plattformvielfalt vom kleinen 386er bis zur großen Multiprozessor-UNIX-Maschine hat er innerhalb kürzester Zeit mit dem höchsten Anteil aller Web-Server viele Konkurrenten auf die hinteren Plätze verwiesen. Das Konkurrenzprodukt zu Apache ist der IIS. Der Apache kann relativ schnell um verschiedene Skriptsprachen wie PHP oder Perl erweitert werden. Es besteht so die einfache Möglichkeit auch dynamische Webseiten zu erzeugen.

6.2 Javascript

Am Anfang des Internets bestand das Internet nur aus HTML-Dokumenten, die von einem Browser, z. B. dem 'Netscape Navigator' oder dem 'Internet Explorer', dargestellt werden konnten. Die einzige Möglichkeit eine Interaktivität mit dem Benutzer herstellen zu können, bestand in der Möglichkeit mit einem CGI-Skript dynamische Seiten zu erstellen. Dazu musste aber jedes Mal eine Anfrage an den Server gestellt werden, deren Ergebnis dann von dem Browser dargestellt werden muss. Dieser Prozess benötigt auf der einen Seite Bandbreite, die der Server zur Verfügung stellen muss, und es kann relativ lange dauern, wenn zur selben Zeit viele Anfragen an den Server gestellt werden, da dieses Skript für jede Anfrage aufgerufen werden muss. Mit der Entwicklung von Javascript wurde dieser Mangel beseitigt, da es von nun möglich war die Darstellung eines angeforderten HTML-Dokumentes auch auf der Client-Seite zu ändern. Das Prinzip der Programmiersprache besteht darin, dass jedes dargestellte Element der Seite als ein Objekt aufgefasst wird, das mittels zugehöriger Methoden und Eigenschaften analysiert und verändert werden kann. So ist es möglich Formularfelder zu verändern oder hinzuzufügen.

Außerdem kann auf Events, wie das Beenden des Ladens der Seite reagiert werden. Aber auch auf Aktionen des Benutzers, wie das Anklicken eines Links oder eines Bildes, kann durch z.B. das Austauschen eines Bildes reagiert werden. Die Darstellung der Seite erfolgt bedeutend schneller, da es nicht mehr nötig ist, mit dem Server zu kommunizieren, sondern die Auswertung erfolgt lokal auf dem Rechner des Clients.

6.3 Linux

ist eine frei kopierbare Variante des Betriebssystems UNIX, die als Alternative auf IBM-kompatiblen PCs ab 80386- Prozessor läuft. 1991 begann der finnische Student Linus Torvalds mit der Entwicklung. Die erste öffentliche Linux-Version 0.01 erschien am 17.09.1991 (500 KByte Quelltexte, verteilt auf 10.000 Codezeilen). Seit etwa 1992 wird Linux von einer Reihe von Programmierern in Zusammenarbeit mit Torvalds weiterentwickelt. Inzwischen programmieren rund um die Welt Tausende daran, da die Quellcodes gleich mitgeliefert werden. Im März 1994 gab Torvalds die Version 1.0 frei. Das Betriebssystem ist kostenlos und sehr stabil. Es gibt bereits zahlreiche Software-Distributionen für Linux. Linux enthält alle Bestandteile für einen Web -Server. Es organisiert sehr zuverlässig die User-Rechte.

Linux ist neben dem Apache-Server eines der Vorzeigeprodukte der Open Source - Bewegung. Der Siegeszug begann, als Linus Torvalds am 09.06.1996 Kernel 2.0 auf die offiziellen FTP -Server stellte.

6.4 Mac OS X

Im Jahre 2001 wurde von Apple ein auf BSD Unix basierendes Nachfolgebetriebssystem für das in die Jahre gekommene Mac OS 9.x vorgestellt. Mac OS X verbindet als erstes Betriebssystem die Vorteile einer ausgefeilten grafischen Benutzeroberfläche mit den Stärken eines UNIX-Betriebsystems. Es bietet für das Netzwerk Funktionen, wie WebServer, MySQL-Server und für den einzelnen Benutzer eine gute 'Office-Fähigkeit'. Seine wirklichen Stärken konnte es aber in diesem Projekt ausspielen: Es sind dies die mitgelieferten Entwicklungstools. So können mit Hilfe der mitgelieferten Frameworks und der Entwicklungsumgebungen ohne Probleme Programme entwickelt werden, die über ein einheitliches Benutzerinterface verfügen. Eine ähnliche Strategie wird im Moment mit der .NET-Technologie von Microsoft versucht. Bestehende C-Software kann ohne Probleme in die Programme integriert werden.

6.5 MySQL

MySQL ist eine echte Multi-User, Multi-Treaded SQL-Datenbank und wird von vielen großen Providern oder auch Suchmaschinenbetreibern eingesetzt. MySQL ist eine Client/Server Implementierung, die aus einem Server-Dämon 'mysqld' und vielen Client Programmen, sowie Bibliotheken für die Programmiersprachen PERL, PHP und ASP besteht.

Structured Query Language (SQL) ist eine standardisierte Datenbanksprache, die das Speichern, Updaten und den Zugriff auf Informationen erleichtert. Beispielsweise kann man Produktinformationen eines Kunden auf einem WWW-Server speichern und abrufen. MySQL ist äußerst schnell und flexibel genug, um sogar Bilder und Log-Dateien darin abzulegen. In dieser Arbeit sind alle Datenbanken mit dieser Datenbank angelegt worden. Dabei ist es nie zu irgendwelchen Datenverlusten oder Abstürzen des Server-Dämons gekommen. Für den Bildungsbereich ist die Nutzung der Datenbank kostenlos.

6.6 PHP

PHP ("PHP = Hypertext Preprocessor") ist eine serverseitig interpretierte, in HTML eingebettete Skriptsprache. Die Syntax ist ähnlich zu C, Java und Perl, und erweitert durch PHP-eigene Features wie z.B. Kommandos zur Integration von Datenbanken. PHP gibt dem WWW-Anwendungs-Entwickler ein einfach erlernbares und gleichzeitig mächtiges Werkzeug zur Erstellung von Web-Seiten dynamischen Inhalts an die Hand. PHP existiert sowohl für Unix als auch für die Windows 95/98/NT Plattform. Sehr beliebt ist die Integration von PHP als Modul in den Apache-Webserver, desweiteren ist die Ausführung via CGI möglich. In dieser Arbeit sind alle über das Internet direkt erreichbaren Applikationen in PHP entwickelt worden. Die Sprache bietet den Vorteil, dass für fast alle Programmierungsprobleme Funktionen zur Verfügung stehen. Dies war von großem Vorteil für die Erstellung der Spektren-Darstellungen. Hier konnte einfach auf die Grafik-Funktionen zurückgegriffen werden.