UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS
DEPARTAMENTO DE QUÍMICA E BIOQUÍMICA

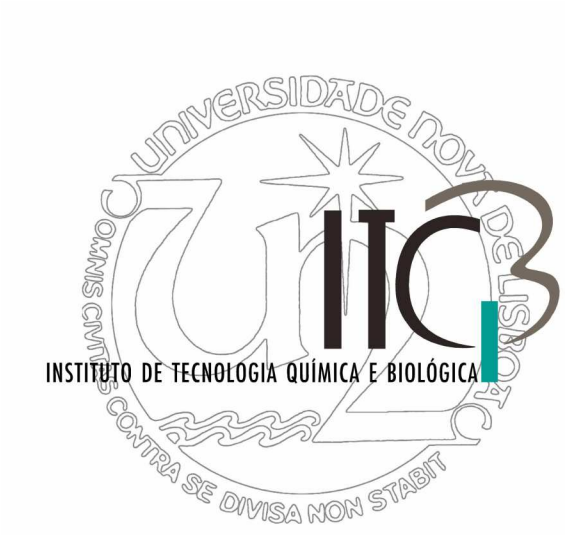
Studying the folding of peptide dendrimers using molecular simulation methods

Luís Carlos Santos Filipe

Tese orientada por Doutor António M. Baptista e Prof. Doutor António Ferreira

2009

Dissertation to attain a Master's degree in Biochemistry from Faculdade de Ciências da Universidade de Lisboa. The work presented here was conducted at Instituto de Tecnologia Química e Biológica da Universidade Nova de Lisboa under the supervision of Doctor António M. Baptista and Doctor António Ferreira



"Proliferam confusões sobre o que é um trabalho de pesquisa original. Não é uma reflexão magna sobre o passado e o futuro do universo, nem é uma nova síntese da filosofia ocidental, de Parménides a Popper. É, habitualmente, uma investigação sobre um tema minúsculo e muito especializado, com conclusões modestas e com impacto reduzido. Mas exige muito trabalho original."

Nuno Crato, *in* Passeio Aleatório

Abstract

Dendrimers are a family of branched compounds that share a common layout where wedges emerge radially from a core by means of a regular branching pattern. Peptide dendrimers are a specific kind of dendrimers formed by alternating functional amino acids with branching diamino acids.

There has been increasing interest in the synthesis of peptide-based dendritic architectures modelling specific aspects of biological function. Some results are already available, demonstrating these molecules ability to act as enzyme models and to mimic natural ligands.

Unfortunately, most studies concerning peptide dendrimers lack structural information at the molecular level. The theoretical study published so far, reported peptide dendrimers presenting shapes close to spheres, though experimental studies on the same systems suggest the existence of more disordered states.

Herein, we characterize five third-generation peptide dendrimers (B1, B1H, B1HH, B1HHH and C1) through multiple long molecular dynamics simulations (MD), and analyse their conformational details and folding preferences in solution. Special emphasis is placed on the analysis of conformational trends representative of the examined models. The conformational sampling results, obtained through MM/MD simulations, were scrutinized using several approaches. Namely, histogram analysis, *phi-psi* dihedral distributions, inter-residue distance matrices, shape analysis and principal coordinate analysis. The adequacy of each approach to discern the conformation space of peptide dendrimers is discussed.

Using these analysis procedures we were able to observe two distinct types of behaviour (*sphere-like* and *bowl-like* structures), both asserting the enormous structural flexibility characterizing these molecules; and the myriad of conformational states available to them.

Our conclusions can be interpreted together with the available experimental results, contributing to a synergistic understanding of the structure-function relation in peptide dendrimers, and casting the bases for novel knowledge-based applications.

Key Words: peptide dendrimers; folding; diffusion coefficient; molecular mechanics; molecular dynamics simulation; gyration radius; gyration space; root mean square deviation; principal coordinate analysis.

Resumo

Os dendrímeros são uma família de compostos ramificados que partilham uma arquitectura comum, onde diferentes cadeias emergem radialmente de um mesmo núcleo (ou centro) através de um padrão de ramificação regular. Os dendrímeros peptídicos, são uma classe particular de dendrímeros, constituída por estruturas que incorporam de forma alternada e iterativa resíduos de amino-ácidos funcionais (resíduos de espaçamento) com resíduos de diamino-ácidos ramificados (resíduos de ramificação). Os resíduos de diamino-ácidos ramificados promovem a bifurcação das cadeias peptídicas e a aquisição da estrutura dendrítica.

A possibilidade de sintetizar dendrímeros com composições que mimetizem as funções de moléculas biológicas, constitui o aliciante para a investigação neste campo cientifico. Em particular, a síntese planeada e controlada de estruturas dendríticas baseadas nos componentes apresentados pelas moléculas biológicas, como peptidos ou glícidos, constitui um desafio atractivo pelas potenciais aplicações que daí podem emergir.

De facto, já foram reportados dendrímeros peptídicos que modelam aspectos específicos de funções biológicas, tais como: modelos enzimáticos para catálise dirigida ("enzimas artificiais"); mimetização de co-factores naturais (de que é exemplo a vitamina B12); transportadores de fármacos, pois quando acopladas aos ligandos adequados estas moléculas têm a capacidade de aderir à membrana celular.

É também importante referir que várias destas moléculas têm sido estudadas enquanto modelos de *folding* das proteínas naturais, pois investigações experimentais

indiciam que alguns dendrímeros peptídicos podem apresentar, em solução, uma estrutura compacta semelhante à das proteínas globulares.

Contudo, a maioria dos estudos experimentais realizados até à data são omissos no que concerne a informação estrutural, e carecem do enquadramento adequado a nível molecular e atómico. O único estudo teórico publicado sobre dendrímeros peptídicos parece confirmar a ideia de que, em solução estas moléculas apresentam, de facto, formas semelhantes a esferas, isto apesar de existirem evidências experimentais que sugerem a existência de estados conformacionais mais desordenados, nesses mesmos sistemas.

Considerando o grande interesse que estas moléculas têm vindo a despertar, a verdade é que pouco se sabe sobre o seu arranjo estrutural tridimensional, e sobre os processos que a ele conduzem (*folding*). Nesta tese tentamos preencher algumas destas lacunas.

Para tal, procedemos à caracterização de cinco dendrímeros peptídicos de terceira geração (que designamos por B1, B1H, B1HH, B1HHH e C1) com diferentes constituintes peptídicos.

Os sistemas que escolhemos como objecto de estudo, estão directamente relacionados com a coordenação da aquocobalamina (análogo da vitamina B12) a dendrímeros peptídicos, ainda que apenas três deles tenham sido sintetizados e caracterizados experimentalmente (B1, B1H e C1). Deste modo, pretendemos não só investigar as suas preferências conformacionais, mas também inferir possíveis relações entre a sua estrutura e a capacidade para desempenhar uma função análoga à das moléculas biológicas (transcobalamina).

É importante salientar que de entre os dendrímeros que foram sintetizados experimentalmente, e que são também aqui estudados, os que apresentam maior capacidade de coordenação com a aquacobalamina, são os que possuem um menor número de resíduos com potencial de coordenação. Este aparente paradoxo é por si só interessante e pode estar interligada com aspectos mais estruturais.

Como temos por objectivo compreender as alterações e a variabilidade subjacentes às estruturas tridimensionais dos diferentes dendrímeros, empregamos metodologias adequadas ao detalhe da escala que pretendemos investigar. Nomeadamente, métodos computacionais de simulação molecular (MM/MD). Optámos portanto por simular

cada um destes cinco sistemas através de múltiplas e longas simulações de dinâmica molecular, utilizando a água enquanto solvente explícito. Com efeito, no trabalho que conduziu a esta tese, realizamos simulações que contabilizam aproximadamente 1 μs^{-1} para cada um dos dendrímeros em estudo. No que respeita a estes sistemas, isto é muito superior ao tempo simulado em estudos anteriores.

Nas últimas décadas a investigação científica tem beneficiado imenso do avanço das técnicas de simulação computacional, que providenciam resultados e formas de escrutinar sistemas, que são de outra forma normalmente inacessíveis.

A dinâmica molecular, especificamente, permite "seguir" a evolução temporal dos átomos que constituem um sistema, através da integração das equações de Newton para o movimento de corpos. É inclusive um dos métodos computacionais de eleição para estudar fenómenos biomoleculares.

Os resultados obtidos com esta técnica de amostragem conformacional permitiram-nos analisar e identificar de forma adequada, os detalhes estruturais de cada um dos dendrímeros peptídicos. Colocámos especial ênfase nos arranjos estruturais mais estáveis. As conformações tridimensionais obtidas a partir das trajectórias resultantes das simulações, foram agrupadas de forma a obtermos os *ensembles* conformacionais característicos de cada dendrímero.

Sobre estes conjuntos de conformações realizamos várias análises. Começámos por investigar algumas das propriedades que caracterizam estes sistemas, como o raio de giração, o número total de ligações de hidrogénio, a distância máxima entre os dois átomos mais afastados de cada estrutura, a superfície acessível ao solvente, entre outros. O raio de giração revelou ser a propriedade que individualmente, melhor espelha as variações intrínsecas a estes sistemas.

Adicionalmente, procedemos também à caracterização da distribuição dos valores de *phi-psi* característicos dos diedros de cada um dos dendrímeros. Complementámos esta análise com o estudo das matrizes que reflectem as distâncias mínimas entre os resíduos de todas as conformações.

Posteriormente aplicámos metodologias de análise conformacional que envolvem a determinação da energia livre, associada a diferentes coordenadas reaccionais (ou de *folding*) para cada estrutura nos diferentes *ensembles*, obtendo assim as correspondentes superfícies energéticas (*folding landscapes*). Utilizamos esta

abordagem por forma a obter *folding landscapes* bi- e tridimensionais. Em especifico, utilizamos como coordenadas de *folding* os valores do raio de giração, do *root mean square deviation* (RMSD), dos componentes principais do tensor do raio de giração diagonalizado, e os valores para a posição relativa das diferentes conformações, num espaço concordante com a matriz de RMSD, utilizando para tal o método de análise das coordenadas principais (PCoorA).

Utilizando o tensor do raio de giração, foi possível investigar a forma dos arranjos estruturais de cada dendrímero peptídico, tendo inclusive sido definido um espaço tridimensional baseado nos componentes principais do tensor diagonalizado (espaço de giração).

A capacidade de cada uma destas abordagens para discriminar de forma adequada o espaço das conformações dos dendrímeros peptídicos é discutida ao longo da tese.

Dos diversos procedimentos de análise conformacional empregues, resulta uma clara indicação de que, em solução, os dendrímeros peptídicos podem apresentar dois comportamentos preferenciais distintos: estruturas compactas que privilegiam as interacções entre os diferentes resíduos, semelhantes a esferas (*sphere-like*); e estruturas "abertas" com as diferentes ramificações espaçadas, em que as interacções entre resíduos não adjacentes são minimizadas, semelhantes a taças (*bowl-like*). Ambas estas configurações atómicas consubstanciam a enorme flexibilidade estrutural que parece caracterizar estas moléculas, dando provas da miríade de estados conformacionais que lhes estão acessíveis.

Foi ainda possível verificar a existência de evidências que suportam a ideia de que estas moléculas possuem uma grande robustez estrutural. Isto é, pequenas alterações na composição dos resíduos de amino-ácidos que as constituem não parecem desencadear alterações conformacionais significativas nos arranjos estruturais preferenciais.

Através da comparação entre o coeficiente de difusão experimental disponível para um dos dendrímeros, e o coeficiente de difusão calculado com base nas trajectórias obtidas por simulação, foi possível verificar que os modelos utilizados, reflectem de forma adequada os sistemas experimentais. Concluiu-se também que o campo de força (*force field*) GROMOS 53A6 possui a capacidade de transferabilidade apropriada para lidar com estas moléculas. Os pontos fortes e fracos dos nossos

modelos são discutidos ao longo da tese. Durante este trabalho foi ainda desenvolvida e implementada uma metodologia que permite o cálculo eficiente do RMSD entre estruturas dendríticas.

As conclusões apresentadas nesta tese podem ser interpretadas juntamente com os resultados experimentais disponíveis, de forma a contribuir para uma compreensão sinérgica da relação entre a estrutura e a função dos dendrímeros peptídicos, lançando as fundações para aplicações inovadoras.

Acknowledgments

I would like to start this thesis by acknowledging the contributions made by António Baptista and Miguel Machuqueiro to its happening. It is because of them that I could not in conscience write this dissertation in the first-person singular and felt compelled to write it in the first-person plural.

My sincere thanks to António, for "deserting from the laboratory bench" soon enough. His teachings and ideas have helped me to enjoy biochemistry. I thank Miguel for his guidance, motivation and priceless (although sometimes neglected) advices.

I would also like to acknowledge my lab colleagues, Sara and Pedro, for their fellowship, and for the many and helpful discussions.

I would further like to emphasize the role of the people from the Protein Modeling group at ITQB. Their advices, availability and friendship were essential for the completion of this work.

Thanks to Prof. António Ferreira for giving me this opportunity.

I acknowledge doctors Tamis Darbre and Jean-Louis Reymond from the University of Bern, for providing original data before its publication.

I would specially like to thank my parents and my brother, for their unconditional support and patience.

Last, I would like to express my gratitude to the person who most helped me to keep my physical and mental sanity, Rita. By whom I am crazy about and without whom this thesis would have taken twice as long to write.

Contents

List of Abbreviations and Symbols

Amino Acid Residues

Three Letter	One Letter	Amino Acid
Ala	A	Alanine
Arg	R	Arginine
Asn	N	Asparagine
Asp	D	aspartic Acid
Cys	C	Cysteine
Glu	E	glutamic Acid
Gln	Q	Glutamine
Gly	G	glycine
His	H	histidine
Ile	I	isoleucine
Leu	L	leucine
Lys	K	lysine
Met	M	methionine
Phe	F	phenylalanine
Pro	P	proline
Ser	S	serine
Thr	T	threonine
Trp	W	tryptophan
Tyr	Y	tyrosine
Val	V	valine

Abbreviations

3D	three-dimensional
Ace	acetyl group, or corresponding acetyl cap
Amb	4-aminomethyl (benzoic) acid
B_n	branching unit n, where n is a natural number
CPU	central processing unit
Dap	s-2,3-diaminopropanoic acid
EDA	Ethylenediamine
G_n	generation n, where n is a natural number
HB	hydrogen bond
Lyr	branching Lysine

MC	Monte Carlo
MD	molecular dynamics
MM	molecular mechanics
Nh2	amine group, or corresponding anime cap
NMR	nuclear magnetic resonance
NPT	isothermal-isobaric (ensemble)
PAMAM	poly(amidoamine)
PCA	principal component analysis
PCoorA	principal coordinate analysis
PDB	protein data bank
PMF	potential of mean force
PPI	poly(propyleneimine)
RDF	radial distribution function
RMSD	root mean square deviation
SPC	single point charge (water)
X_n	spacer unit n, where n is a natural number

Latin Symbols

A	generic property or quantity A
$\mathbf{a}$	acceleration
B	generic property or quantity B
C_α	alpha carbon
$C_{R_g}(t)$	time autocorrelation function for the radius of gyration
C_{AA}	autocorrelation function of property A
C_{AB}	cross-correlation function of properties A and B
C'_{AA}	non-normalized autocorrelation function
D	diffusion coefficient
$\mathbf{E}$	unit vector
E	total free energy
$\mathbf{F}_i$	force exerted on particle i
G	Gibbs free energy
I	moment of inertia
I_{xx}, I_{yy}, I_{zz}	principal components of inertia tensor
k_B	Boltzmann constant
L_1, L_2, L_3	principal components of the radius of gyration tensor
M	total mass of a set of particles
m_i	mass of particle i
N	total number of particles
N_A	Avogadro number
P	pressure
q_i	value for the point charge of particle i
R	gas constant
$\mathbf{r}$	general position vector
R_g	radius of gyration
R_H	hydrodynamic radius
$\mathbf{r}_i$	position vector of particle i

r_i	position vector length
$\mathbf{r}_{ij}$	vector defined between particles i and j
r_{ij}	norm of the vector defined between particles i and j
s	a single state of the system
S	radius of gyration tensor
T	absolute temperature
t	time
V	potential energy function
$\mathbf{v}$	velocity vector
V_s	volume of state s
z	each molecule in an ensemble
Z	total number of molecules in an ensemble

Greek Symbols

ε_0	electric permitivity of vacuum
δA	deviation of A from ensemble average
$\sigma(A)$	root mean square deviantion for dynamic variable A
$\Delta(p,T)$	partition function in the isothermal-isobaric ensemble

Other notations

$\partial / \partial t$	partial derivative with respect to time
$\nabla_{\mathbf{r}_i}$	gradient with respect to molecular position
$\langle \ldots \rangle_{ens}$	ensemble average
$\langle \ldots \rangle$	average

Scope and Structure

The main objective of this thesis is to address key questions and provide an improved understanding on the behavior of peptide dendrimers in aqueous media. In particular, the focus of this thesis concerns the folding preferences of these molecules.

In the first chapter we introduce the key concepts related with peptide dendrimers, and dendritic molecules in general. The specific subjects and goals of the current work are also presented.

The second chapter contains a global overview of the molecular mechanics/molecular dynamics simulations topics of importance to the present study. Moreover, Chapter 2 encompasses the models and simulation details.

In the third chapter we present the results of our simulations, and the analysis undertaken to explore the conformational preferences of peptide dendrimers. The strengths and weaknesses of the models used are also discussed.

Finally, the fourth chapter summarizes the main aspects investigated in this thesis, and discusses their repercussions and implications.

Chapter 1

Introduction

The present work is intended, and was written, to a broad audience within the scientific community. In particular, it is not assumed that the common reader has a detailed knowledge on the topic of dendrimers in general and peptide dendrimers in particular. Therefore, a brief discussion on both topics is presented as a mean to facilitate an integrated and synergistic understanding of the current work.

1.1. What are dendrimers? Definition and general characteristics

Ideally, dendrimers are a family of perfect nanosized, globular, monodispersed macromolecules characterized by a regular and unique tree-like three-dimension branching architecture and compact spherical geometry in solution (see Figure 1.1(a))[Fréchet 2002]. However, due to the legion of molecular constructions that have been classified as dendrimers in the past years, together with the absence of a

consensual/unified international nomenclature, a broader definition is more suited[1], namely: "Dendrimers are diverse branched compounds with different composition, structure, molecular weight, surface groups, valency, physicochemical properties and biological activities" [Niedrhafner 2005]. From this comprehensive definition one can discern that the feature common to all dendrimers is their global layout. They are composed of individual "wedges" or "dendrons" that emerge radially from a central focal point by means of a regular branching pattern with repeated units, and where each layer of concentric branching units constitutes one complete generation and is identified with a specific generation number (see Figure 1.1(b)). This branching architecture leads to a controlled increase in a dendrimer's molecular weight, size, and number of surface groups [Fréchet 2002; Medina 2009]. Because dendrimers are built from $(AB)_n$-type monomers (with A representing a spacer unit and B a branching unit), each layer or "generation" of branching units doubles or triples (n = 2 or 3) the number of peripheral functional groups [Cloninger 2002]. Therefore these structures are characterized by a combination of high end-group functionality and a compact, precisely defined molecular structure containing three topologically different regions (core, branches, and surface), each of which can exhibit functional properties modulated by the dendrimer as a whole [Medina 2009].

As previously mentioned, the dendritic structure is characterized by "layers" between each focal point (or cascade) called generations (shown in different colors on Figure 1.1(b)), in spite of that, the exact numbering of generations has been the subject of some confusion [Boas 2004]. In the present thesis, and following the definition employed by other authors [Boas 2004; Niedrhafner 2005], the dendrimer generation is defined as the number of focal points (or cascade points) when going from the core to the surface; a generation 5 (G5) dendrimer has thus 5 cascade points between the core and the surface. The core is sometimes denoted "generation zero" (G0), as no cascade points are present (Figure 1.1).

[1] Normally the definition considered will depend on the author and will be particularly influenced by its scientific field.

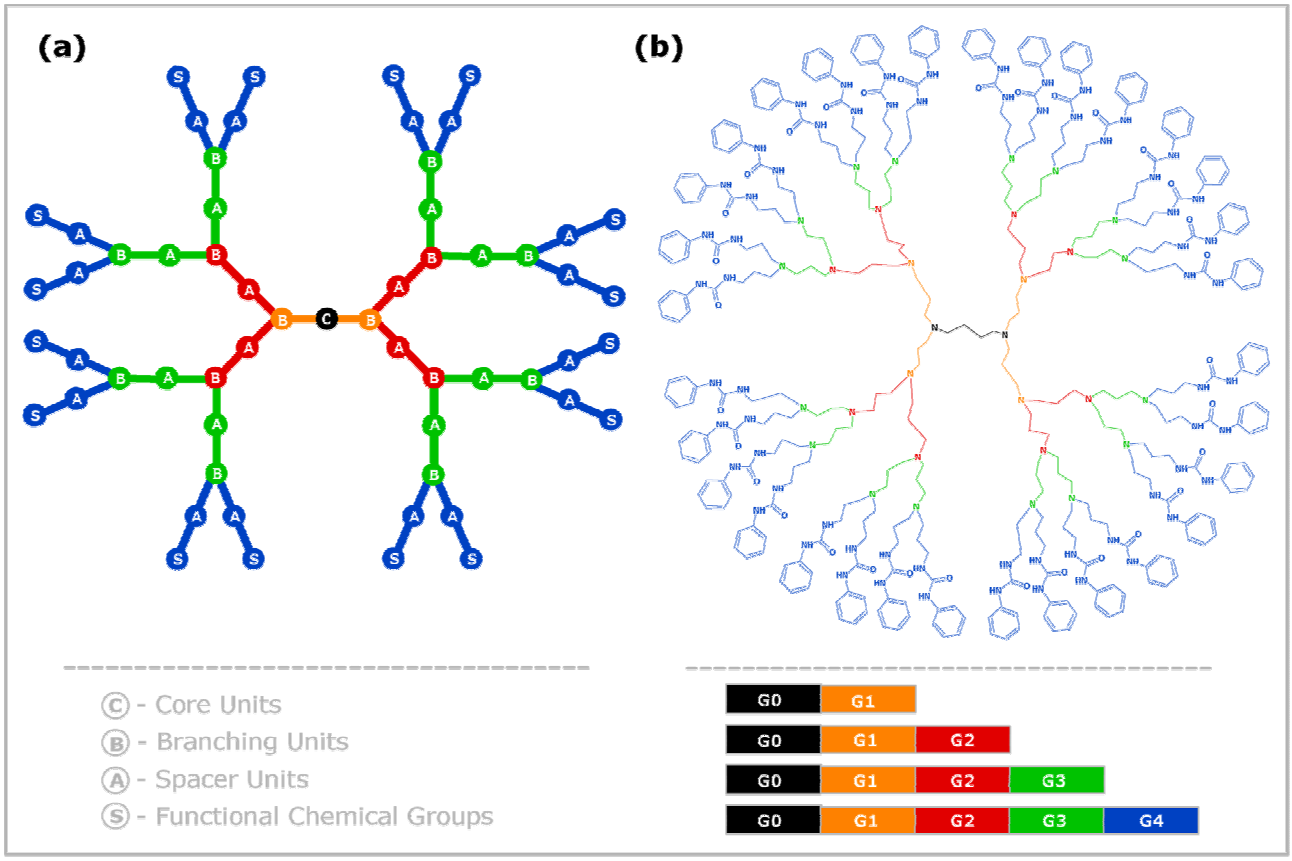


Figure 1.1 – Schematic representation and chemical structure of a Poly(propylene amine) dendrimer

(a) Schematic representation of a trifuncional fourth generation dendrimer. Different generations are highlighted trough the use of specific colors, and different constituents of the dendritic structure are emphasized through the use of a one letter code. The represented dendrimer is built using AB_n-type monomers, therefore each new layer (or "generation") of branching units doubles the number of peripheral functional groups. **(b)** Chemical structure of a fourth generation poly(propylene amine) dendrimer, containing sixteen N-phenylamide units in its periphery.

Not all regularly branched molecules are "ideal" dendrimers because properties of the dendritic state, such as core encapsulation and unusually low intrinsic viscosity in solution, are reached only when globularity is achieved at a certain generation or size threshold [Ballauff 2004; Hawker 1993(a); Hecht 2001]. Still, dendrimers come in many formats presenting a myriad of shapes and sizes, and while a few present themselves as rigid molecules, many others have the ability to modify the conformational arrangement of their various constituting blocks to minimize the overall free energy [Fréchet 2002].

Since the seminal works in this field [Buhleier 1978; Tomalia 1985; Newkome 1985], a large number of dendrimer structures have been developed and have become a subject of intense interdisciplinary research efforts.

Dendrimers have been the topic of intensive research and study, and the so-called "dendrimer chemistry" has become a rapidly expanding research area at the interface of traditional synthetic organic chemistry and polymer science. Details such as the nature of the building blocks, the composition of the branching units, the effect of the number of generations or pH value [Maiti 2005; Maiti 2004; Klos 2009], among others, have aroused scientific curiosity in experimental an theoretical fields. The "dendrimer subject" is in our days a multidisciplinary area ranging from physics to biomedical sciences.

Dendrimers are produced in an iterative sequence of controlled reaction steps, in which each additional iteration leads to a higher generation material (see below on the topic concerning – Dendrimer Synthesis). The first example of an iterative synthetic procedure toward well-defined branched structures was reported by Vögtle [Buhleier 1978], who named this procedure a "cascade synthesis". Still the first dendritic structures that were thoroughly investigated and that received widespread attention are Tomalia's Poly(amidoamine) – PAMAM – dendrimers [Tomalia 1985], and Newkome's "arborol" systems [Newkome 1985].

Since then a large variety of dendrimer building blocks have been reported, including carbohydrates and amino acids (see below); also a great number of terminal groups have been modified with suitable functional end groups. Figure 1.1(b) shows a typical example of a fourth generation dendritic structures with defined end groups [Ballauff 2004]. As a consequence it comes without surprise that these distinct and unique structure with characteristic mechanical properties, represent an interesting topic for theoretical research in addition to the creation of innovative applications.

Many of the intriguing properties of dendrimers as well as their syntheses and possible applications are discussed in excellent books and reviews that have been published by various experts in the field and that might complement the information provided in this thesis [Medina 2009; Ballauff 2004; Fréchet 2002; Dykes 2001; Bosman 1999; Boas 2004].

1.2. Brief Dendrimer History

Dendritic architecture is perhaps one of the most universal topologies observed on our planet. Innumerable examples of these patterns may be found in biological systems such as tree branching, tree roots, plant and animal vasculatory systems, neuron networks, dendritic cell, among others [Tomalia 2002].

One might speculate that these are evolutionary architectures that have been optimized to provide structures manifesting maximum surface exposure for optimum exploitation of the topological potential of a system/molecule [Tomalia 2002].

Although perhaps first conceptualized by Flory [Flory 1941], the synthesis of multi-branched compounds was only pioneered by Vögtle and coworkers in 1978 with the so-called "cascade molecules", that were obtained by a repetitive reaction of mono- and diamines with a central core [Friedhofen 2006]. At that time it required a significant shift in polymerization strategies and synthesis methodologies.

The field was further developed in the mid-1980s when Newkome [Newkome 1985] synthesized tree-like molecules termed "arborols" and Tomalia [Tomalia 1985] reported the synthesis of poly(amindoamide) and coined the term "dendrimer" for this class of compounds.

Tomalia and his team described in their published paper the iterative coupling of ethylene diamine to a central ammonia core to produce a series of branched macromolecules named "starburst dendrimers" [Tomalia 1985]. The word dendrimer is derived from the Greek word "dendron", which means "tree" or "branch", and "meros" meaning "part of".

These early dendrimers were prepared in an iterative sequence of steps to achieve growth and branching, starting from a central core unit. This approach would later be termed "divergent". However, a shift occurred in the early 1990s, when Hawker and coworkers constructed dendritic materials from the periphery inward [Hawker 1990]. This "convergent" approach was first demonstrated with poly(ether) dendrimers, and provided access to dendrimers of unprecedented purity allowing also greater control

over the placement of functional groups in the macromolecule than could be achieved with divergent methods [Sadler 2002].

Most of the work done with dendrimers since then was motivated by the idea that dendrimers possess a hollow core and a dense shell. The underlying assumption was that the segmental density increases from the center to the periphery, and thus it should be possible to use such molecules as carriers ("dendritic boxes") [Ballauff 2004]. This dense-shell picture has been strongly supported by a paper wrote by de Gennes and Hervet [Gennes 1983] in 1983 in which they presented the first theoretical treatment of dendrimers and concluded that the density increases towards the periphery of the molecule. However, Lescanec and Muthukumar [Lescanec 1990] have latter shown that a dendritic structure made up from flexible bonds should exhibit its maximum density at the center of the molecule (dense-core model). This pioneering work has been the starting point of a considerable number of theoretical studies [Klos 2009; Breslow 2007].

A considerable number of molecular simulations have been dedicated to dendrimers in the last few years, but the first simulation work on dendrimers was performed by Naylor and coworkers in 1989 [Naylor 1989].

They carried out atomic-scale MD simulations of balanine starburst dendrimers up to the seventh generation by using atomic-level force fields.

The following table reflects the present authors choice from the available literature on what we consider to be the most relevant articles on the dendrimer field.

Table 1.1 – Main articles in the field of dendrimer chemistry

This table presents the publications that the authors consider as milestones in the study and development of dendrimers.

Title	Reference	Year	Relevance
"Cascade"- and "Nonskid-Chain-like" Syntheses of Molecular Cavity Topologies	[Buhleier 1978]	1978	First synthesis of multi-branched compounds
Statistics of starburst polymers	[Gennes 1983]	1983	First theoretical treatment of dendrimers
A New Class of Polymers: Starburst-Dendritic Macromolecules	[Tomalia 1985]	1985	Synthesis of the first poly(amindoamide) compounds; the word ”dendrimer” was employed for the first time; development of the divergent approach to dendrimer synthesis
Cascade Molecules: A New Approach to Micelles	[Newkome 1985]	1985	Creation of “arborol” systems
Starburst Dendrimers.5. Molecular Shape Control	[Naylor 1989]	1989	First molecular simulation work of dendrimers
Preparation of Polymers with Controlled Molecular Architecture. A New Convergent Approach to Dendritic Macromolecules	[Hawker 1990]	1990	First use of the convergent approach synthesis
Solvatochromism as a Probe of the Microenvironment in Dendritic Polyethers: Transition from an Extended to a GlobuIar Structure	[Hawker 1993(a)]	1993	First proof of the existence of distinct microenvironments inside dendrimers
Encapsulation of Guest Molecules into a Dendritic Box	[Jansen 1994]	1994	First report on the encapsulation of guest molecules into a dendritic architecture, the so-called “dendritic box”
Dendrizymes: Expanded ligands for enantioselective catalysis	[Brunner 1995]	1995	First dendritic catalysts
Catalytic Peptide Dendrimers	[Esposito 2003]	2003	First catalytic peptide dendrimers
Molecular Dynamics and Docking Studies of Single Site Esterase Peptide Dendrimers	[Javor 2009]	2009	First molecular dynamic simulations with dendrimers presenting lysine or s-2,3-diaminopropionic acid as branching units

1.3. Different dendrimers with the same architecture – dendrimer families

A variety of molecules can be used as repeating units of dendrimers to diversify the size, shape and viscosity of the dendritic architecture. The most common dendrimers are polyamidoamine (PAMAM) dendrimers and the poly(propylene imine) dendrimers (PPI) [Hawker 1990]. Additionally, many other types of scientifically interesting dendritic systems have been developed, and thus, a variety of dendritic scaffolds have become accessible with defined nanoscopic dimensions and discrete numbers of functional end groups [Bosman 1999].

A brief review on the existing families is in order. Even so, for the eager reader a more comprehensive review on the subject of dendrimer families can be found in some excellent articles [Dykes 2001; Medina 2009; Crespo 2005; Niedrhafner 2005].

PAMAM Dendrimers - Poly(amidoamine) (PAMAM) dendrimers were the first synthesized and commercially available dendrimer family (Figure 1.2(a)) [Lin 2005].

The synthesis of PAMAM dendrimers is initiated using an alkyldiamine core (e.g., ethylene diamine), which reacts via Michael addition with methyl acrylate monomers to produce a branched intermediate that can be transformed to the smallest generation of PAMAM dendrimers. Synthesis of higher generations of PAMAM dendrimers is achieved by sequential Michael addition of methyl acrylate monomers followed by an exhaustive amidation reaction [Medina 2009]. However, dendrimers growth eventually reaches a critical point where the steric crowding of the branching arms limits their development in a phenomenon known as the de Gennes dense-packing effect [Medina 2009; Maiti 2004; Boas 2004].

PAMAM dendrimers (Figure 1.2(a)) prepared by the divergent growth approach of Tomalia *et al.* [Tomalia 1985], are one of the most widely used dendrimer scaffolds in biology [Gillies 2005].

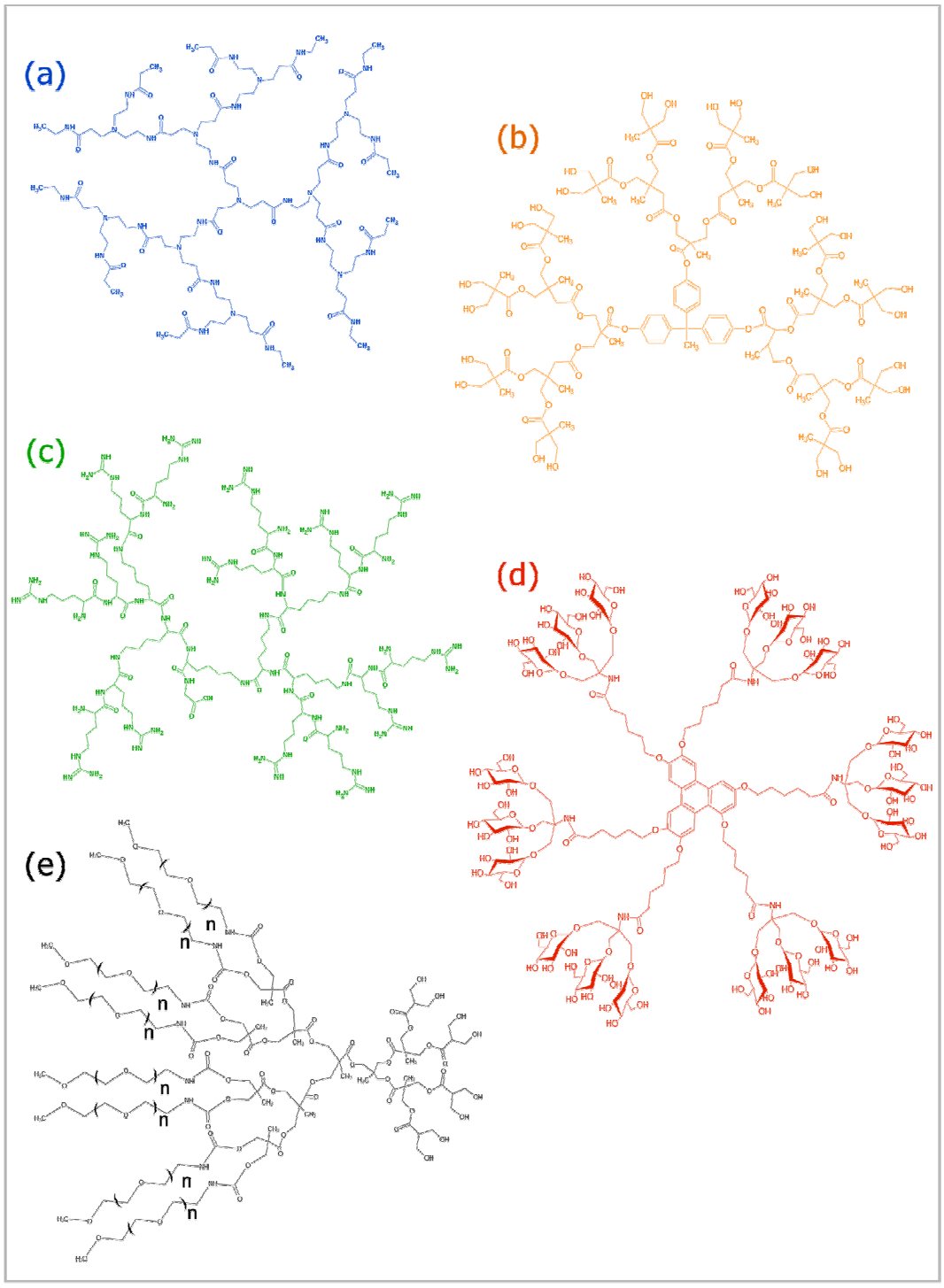


Figure 1.2 – Representative examples of different dendrimer families

Chemical structure of several dendrimers: **(a)** poly(amidoamine) dendrimer – *PAMAM dendrimers*; **(b)** polyester dendrimer based on 2,2-bis(hydroxymethyl)propionic acid – *Biodegradable dendrimers*; **(c)** Arginine-Lysine dendrimer – *Amino acid-based dendrimers*; **(d)** triphenylene-centered glycodendrimer presenting α-D-Allopyranose on its surface – *Glycodendrimers*; **(e)** poly(ethylene oxide) dendrimer – *Asymmetric dendrimer*.

Biodegradable Dendrimers - The need for biodegradable dendrimers emerged as a strategy to produce the desired large molecular weight carriers that achieve high accumulation and retention in tumor tissue [Patri 2002; Shi 2009; Darbre 2006]. Biodegradable dendrimers are commonly prepared by inclusion of ester groups in the dendrimer's backbone, which will be chemically hydrolyzed and/or enzymatically cleaved by esterases in physiological solutions [Medina 2009] (see Figure 1.2(b)). Various polyester dendrimers incorporating monomers such as glycerol, succinic acid, phenylalanine and lactic acid have been prepared by Grinstaff and coworkers [Grinstaff 2002], and their potential use in tissue engineering has been tightly demonstrated.

Amino Acid-Based Dendrimers - Of the many monomer structures possible, those that contain natural or unnatural amino acids (Figure 1.2(c)) are particularly appealing because they are chiral and have the potential to produce dendrimers with enhanced biocompatibility and diversity [Kim 1999; Crespo 2005]. Furthermore, defined three-dimensional structures might be attained through specific folding of the constituent amino acid units [Javor 2008; Breslow 2007]. Optically active protein-mimetic dendrimers have already been synthesized using a library of amino acids including tryptophane, phenylalanine, glutamic acid, aspartic acid, leucine, valine, glycine, and alanine [Kono 2008; Ranganathan 1997; Kim 1999].

Amino acid-based dendrimers are synthesized using one of the following strategies: (1) amino acid or peptide grafting and display on the periphery of a conventional dendrimer or (2) attachment of amino acids or peptides to an organic or peptide core [Sadler 2002; Lockman 2005].

Further details on dendrimers containing amino acid constituents, will be discussed in the section about peptide dendrimers (see below).

Glycodendrimers - The term “glycodendrimer” is used to describe dendrimers that incorporate carbohydrates into their structures. Two reviews nicely summarize the different approaches to glycodendrimer synthesis and other relevant aspects [Turnbull 2002; Bezouska 2002].

The vast majority of glycodendrimers (Figure 1.2(d)) have saccharide residues on their outer periphery, but glycodendrimers containing a sugar unit as the central core from which all branch points emanate as well as glycodendrimers with carbohydrates as the main dendrimer building blocks have also been described [Bezouska 2002; Turnbull 2002]. Glycodendrimers with periphery carbohydrate units have been used to study the protein carbohydrate interactions that are implicated in many intercellular recognition events [Johansson 2008; Veprek 1999(p)]. The so-called "sugar-coating" of dendrimers serve as multivalent recognition structures for sugar-binding proteins, such as lectins [Bezouska 2002; Johansson 2008].

Hydrophobic Dendrimers - The inclusion of hydrophobic regions in the dendrimer structure allows for better encapsulation and efficient solubilization of hydrophobic drug molecules within the dendrimer voids [Kono 2008]. Specifically, dendrimers with hydrophobic cores proved to effectively retain hydrophobic drug molecules in the voids of their branching architecture, mimicking amphiphilic polymer micelles [Fréchet 2002; Patri 2002].

Several groups capitalized on this concept and developed unimolecular micelles using dendrimers with hydrophobic interiors and a hydrophilic surface, which were used to solubilize and encapsulate a number of hydrophobic guest molecules [Zeng 1997; Hawker 1993(b)].

Asymmetric Dendrimers – Imparting asymmetry to dendrimer's architecture can provide a range of novel structures, which may favorably affect their pharmacokinetic profile in vivo. Asymmetric dendrimers are synthesized by coupling dendrons of different generations to a linear core, which yields a branched dendrimer with a nonuniform architecture (see Figure 1.2(e)). Although these are not "ideal" dendrimers, due to the lack of a perfectly symmetric structure, asymmetry allows for tunable structures and molecular weights, with precise control over the number of functional groups available on each dendron for attachment of drugs, imaging agents, and other therapeutic moieties [Medina 2009]. The most recognized asymmetric dendrimers were synthesized by Fréchet and coworkers and are known as the "bow-tie" polyester dendrimers [Gillies 2002; Gillies 2005].

1.4. Strategies for Dendrimer Synthesis

It is worthwhile to review the factors that affect dendrimer design and architecture. These highly branched macromolecules are synthesized by a step-wise approach with either linear or branched building blocks. Unlike most traditional polymers, one can take precise structural control over the molecular weight as well as chemical and physical properties of dendritic macromolecules. Typically, dendrimer synthesis includes the following steps: (1) selection of a suitable initiator that can be converted into a reactive initiator core with good yield; (2) definition of an iterative reaction sequence whereby the reactive initiator core is exposed to appropriate reagents or other reactive (partially protected) branched molecules, thus leading to high yield conversions of branched assemblies with specific molecular surfaces; (3) reiteration of these step-growth or chain-growth sequences to produce dendrimers possessing concentric generations of repeating units and branch junctures [Fréchet 2002].

The terms divergent and convergent ("outward" versus "inward" tier growth, respectively) are used to describe dendritic synthetic strategies [Fréchet 2002].

The two methodologies are complementary and neither is generally better, the choice of the synthetic approach being usually justified by the features desired for the target molecule, the chemistry available for growth, and the specific building blocks used in the construction of the dendritic framework [Fréchet 2002].

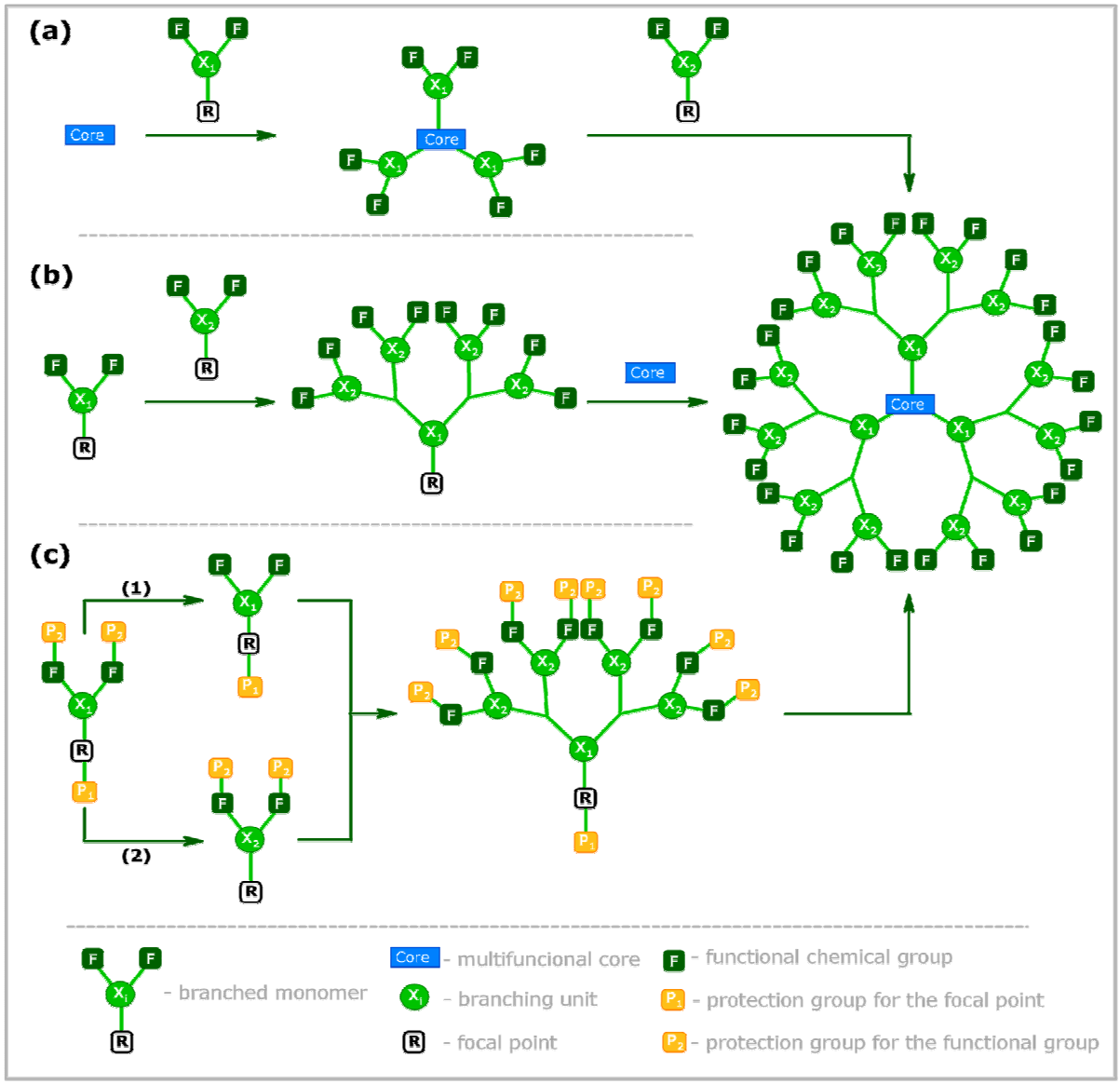


Figure 1.3 – Schematic drawing showing the different methods for synthesis of dendrimers

(a) Divergent synthesis – The divergent dendrimer synthesis is a technique that effectively grows the dendrimer structure from the initiator core to the periphery in a stepwise fashion by iterative addition of monomer units. **(b)**Convergent synthesis – The dendrimer surface is formed by reaction of the chemically activated focal points of branched monomers with the functional groups of another monomer. **(c)**Convergent-Divergent synthesis. See the text for further details on dendrimer synthesis methods.

The divergent dendrimer synthesis is a technique that effectively grows the dendrimer structure from the initiator core to the periphery in a stepwise fashion by iterative addition of monomer units. Specifically, divergent synthesis is initiated by coupling of a monomer unit to a multifunctional initiator core where the dendrimer

generation increases by successive addition of the building blocks to the periphery of the parent dendrimer [Medina 2009] (see Figure 1.3(a)).

However, divergent synthesis has its own limitations besides nonideal growth events, including the difficulty in purifying the final product from structurally similar byproducts and the lengthy multistep reactions, which led to a number of optimizations [Medina 2009].

In 1990, Hawker and Fréchet collaboratively reported a convergent approach to dendrimer synthesis [Hawker 1990] that uses the symmetrical nature of these structures to its advantage and was developed to address the deficiencies of the divergent method. Convergent synthesis begins with the dendrimer periphery units coupled to additional building blocks to form the branching structure, thus constructing dendrons from the periphery toward the central focal point (see Figure 1.3(b)). Each dendron is then coupled through its focal point to a multifunctional core to produce the complete dendrimer.

Unlike divergent synthesis, convergent reaction products are simple to purify since the desired dendrons are substantially different from the reaction byproducts, thus eliminating the need for highly efficient reactions. More importantly, the convergent growth allows unparalleled control over functionality at specified locations of the growing macromolecule, and provides access to numerous novel architectures through the attachment of dendrons to other molecules.

Later on, Fréchet proposed what is called a “double-staged” approach (see Figure 1.3(c)) to increase synthetic yields of higher generation dendrimers by coupling dendrons prepared by convergent synthesis to a dendrimer “hypercore” or a prefabricated lower-generation dendrimer utilized as a multifunctional core [Medina 2009; Liu 1999].

This method allows the use of a dendrimer hypercore that is constructed from flexible units to reduce the steric hindrance during the coupling of the dendron focal points [Wooley 1991].

Although the majority of the dendrimers prepared to date have been built of covalent bonds, many non-covalent dendrimers have also been prepared by a variety of self-assembly processes involving, for example, hydrogen bonding or

supramolecular coordination chemistry [Narayanan 1998; Diederich 2002; Fréchet 2002].

1.5. Shape and Conformational Preferences of Dendrimers

For all the so far studied dendrimers an increase in the generation number results in a marked change on the molecular properties [Ballauff 2004; Han 2005; Maiti 2004] and there has been significant controversy about the shape of dendrimers, and the placement of their chain ends (particularly their functional groups) either at the "periphery" of the macromolecule (dense-shell model) or back-folded within its building blocks (dense-core model). Indeed, a variety of calculations and measurements have suggested back-folding of the chain ends whereas others have ascertained their peripheral arrangement [Ballauff 2004; Naylor 1989; Lescanec 1990; Maiti 2005].

Many dendrimers have been shown to be flexible, whereas some seem to be fairly rigid [Han 2005; Karatasos 2001]. Broadly one can consider that "true" dendrimers, such as those that exhibit an unusual intrinsic viscosity to molecular weight relationship [Tomalia 2002], are globular macromolecules that accomplish considerable rigidity only when a certain number of generations is achieved [Ballauff 2004; Maiti 2004].

Some controversy has also surrounded the issue of the existence of a "cavity" within the dendrimers promoted by the architecture adopted by the different dendritic "arms" [Hawker 1993(a)]. Indeed, several groups have encapsulated numerous different molecules within dendrimers using non-covalent approaches, but this does not means that dendrimers have a permanent and rigid cavity [Gorman 2001].

In fact, a survey of the literature dealing with molecule interactions in dendritic hosts indicates that dendrimers are not only capable of being freely permeated by solvent molecules, but they are also able to rearrange themselves through a significant volume collapse when solvent is removed [Fréchet 2002; Boas 2004].

Overall, and as might be expected, the major contribution to a dendrimer shape and structure are the interactions with the surrounding environment (namely the existing solvent) in order to minimize its free energy [Ballauff 2004].

Depending on the surrounding medium, a dendrimer can present itself as a high volume, fully extended structure with a virtually spherical spatial arrangement, or as a highly flexible macromolecular structure with variously collapsed regions and fluctuating monomer groups, being the final shape determined by the interactions of its various components (core, building blocks, chain-ends), and the interactions with its near enviroment [Fréchet 2002; Klos 2009].

Nevertheless, insights into dendrimers internal structure have been obtained by various experimental techniques (for example spectroscopic, microscopic and viscosimetric studies), clarifying the possible back-folding of the flexible branches [Lescanec 1990], the detailed distribution profile of the monomers [Han 2005] and the structural response to change in different features of the medium (for instance polarity and pH)[Ballauff 2004; Bosman 1999].

1.6. Dendrimer Applications

Research on dendrimers is a part of modern nanoscience with the aim of tailoring material properties at the molecular level [Crespo 2005]. The motivation for the investigations in course rests on what is thought to be the myriad of spatial arrangements accessible to the dendritic architecture, which leads to the possibility of adapting and iteratively modifying these structures to address several practical, specifically biomedical, needs and applications.

There has been an increasing interest in the development of dendritic architectures which model specific aspects of biological function [Smith 1998]. As such, research has increased on the design and synthesis of biocompatible dendrimers and their application to many areas of biological science, presenting practical solutions to areas including drug delivery, magnetic resonance imaging, immunology, development of vaccines, synthetic macromolecules catalysis, protein models, antimicrobials,

antivirals and anticancer therapeutics among others [Gillies 2005; Stiriba 2002; Niedrhafner 2005; Boas 2004].

A comparison of the features of dendrimers with those of linear polymers shows that the dendritic architecture can provide several advantages for drug delivery applications [Liu 1999]. For example, the controlled multivalency of dendrimers can be used to attach several drug molecules, targeting groups and solubilizing groups to the periphery of the dendrimers in a well-defined manner [Gillies 2005]. Dendrimers have evolved to structures with a large number of surface groups that can be utilized to display a range of biological motifs including peptides, proteins, sugars, and targeting agents while carrying a large therapeutic payload either within the dendrimers voids or on their surface. The high loading capacity of dendrimers renders them highly attractive as carriers for delivery of chemotherapeutic agents into tumor tissue for treatment of cancer [Shi 2009]. PEGylated and non-PEGylated dendrimers proved to encapsulate hydrophobic drug molecules into the hollow voids of their branching architecture, which enhance the aqueous solubility and stability of the encapsulated drug molecules. However, controlling the release kinetics of the encapsulated drug remains a challenging task that depends on the hydrophobicity and size of the drug, the generation number of the dendritic carrier, and the type and extent of modification of the dendrimer's surface [Medina 2009; Liu 1999].

The dendritic shell can produce localized microenvironments analogous to those found at the active sites of enzymes, and dendrimers can now be employed in catalysis and molecular recognition [Smith 1998; Kleij 2000; Brunner 1995; Merý 2006].

Ongoing research is aimed at providing a convincing practical demonstration of the obvious advantages that catalysts containing multiple active sites could possess. However, careful design is required to yield dendritic catalysts with distinct advantages such as facile recycling or cooperativity between catalytic centers [Smith 1998; Darbre 2006].

The "dendrizymes" reported by Brunner were among the first dendritic catalysts to model enzymes, with an organometallic active site buried among the branches [Brunner 1995].

Other dendrimers with encapsulated redox-active metal centers have been reported. Gorman and coworkers prepared dendritic iron-sulfur complexes [Gorman 2001]. Their results indicate a versatile new approach to modulating and optimizing the reduction potential of redox catalysts by controlling the polarity of their environment using dendrimer technology. Attachment of dendritic shells, in particular by convergent growth methods, should find increasing future application in tuning the potential of electrophores for use as redox mediators in electrocatalysis or for performing specific tasks in advanced materials design [Smith 1998].

Dendritically encapsulated metalloporphyrins have also attracted wide interest as models for globular heme proteins, and the dendritic generation-dependence of properties of the metalloporphyrin core has been intensively investigated [Diederich 2002].

All three topologically distinct regions (core, branching shell, and periphery) of dendrimers can associate with suitable substrates, and the first examples of these distinctively different recognition events have emerged more then a decade ago [Stevelmans 1996; Medina 2009; Gillies 2005].

Newkome and coworkers showed early on that water-soluble hydrophobic dendrimers act analogously to micelles and that these “unimolecular micelles” can encapsulate hydrophobic guests among their branches [Newkome 1985]. Most dendrimers are flexible enough to accommodate inclusion guests. Indeed dendrimers are not only capable of being freely permeated by solvent molecules, but they are also able to rearrange themselves through a significant volume collapse when solvent is removed [Fréchet 2002; Boas 2004]. This collapse may leave guest molecules trapped inside the dendrimer, especially if favorable interactions exist, as in some “dendritic micelles” [Hecht 2001], or if the dendrimer structure has been “rigidified” to prevent their escape, as in the “dendritic box” of Meijer and coworkers [Jansen 1994; Hecht 2001].

Meijer and coworkers prepared the fifth generation poly(propyleneimine) (PPI) dendrimer and demonstrated its function as a “dendritic box” capable of retaining substrates trapped during synthesis and preventing them from diffusing outwards [Jansen 1994]. An appropriate dendritic surface can therefore control guest binding

within a cascade molecule, which acts as a mimic for biological container and transport systems such as vesicles.

Prospective uses of this fascinating research include transport and slow-release systems for drug delivery [Patri 2002] and encapsulation of fluorescence markers with biological relevance [Hawker 1993(a)].

1.7. Peptide Dendrimers

Herein we are interested in some specific dendrimers, the peptidic ones. Although some general features of dendritic architecture are of relevance to this work (and as such they have been previously discussed), peptide dendrimers are the main subject. They constitute a particular class of dendrimers with a specific framework and a great number of unstudied/unknown features. Still they "seem" to present most of the common major characteristics of other dendrimers. A description on what is currently known on the subject of peptide dendrimers and some definitions are presented next.

Peptide dendrimers can be broadly defined as any dendrimer that contains peptide bonds. This definition would, in theory, include a dendrimer with an amino acid core, branching units, surface functional groups or any combination of the three as a peptide dendrimer [Niedrhafner 2005] (see Figure 1.2(c) for an example). This definition could be further broadened if we consider all types of amino acids including naturally occurring amino acids as well as the unnatural amino acids that have already been utilized in peptide dendrimer synthesis as both branching units and surface functional groups [Sadler 2002; Veprek 1999(a)].

Nowadays peptide dendrimers vary from low molecular weight species of 2 kDa to large protein-like constructs with more than 100 kDa [Sadler 2002]. The size and complexity of the individual dendrimers are determined by two factors: the number of layers of branching units (generation number) and the surface supporting the terminal functional groups (which can be large peptides or macromolecules of substantial size). Similarly to other dendrimers, to construct such complex and heterogeneous macromolecules, synthesis methods were designed based upon the proper

combination and manipulation of temporary and permanent protecting groups, as well as the choice of efficient coupling and branching reagents for the controlled formation of the peptide bonds [Sadler 2002].

Peptide dendrimers are often broadly divided into three types [Sadler 2002; Niedrhafner 2005]. The first are grafted peptide dendrimers. These are conventional dendrimers with either unnatural amino acids or organic groups as the branching core and peptide or proteins attached as surface functional groups. Of the three types, the grafted peptide dendrimer is the largest in terms of size because they generally contain high generation numbers of branching cores. In contrast, peptide dendrimers of the second type are essentially branching polyamino acids (without spacer residues). Consequently, they tend to be the smallest by size, with the core consisting of natural amino acids and the terminal amino acids acting as peripheral functional groups (see Figure 1.2(c) for examples). The third type, consisting mostly of peptides, have been the ones traditionally known as peptide dendrimers. In this group the core consists of amino acids and the surface functional groups are also peptidyl chains[2] (see Appendix A for examples).

This thesis will focus on some particular peptide dendrimers of the third type. Little is known concerning the solution structural properties of these peptide dendrimers, and different behaviors have been ascertained. The most recent theoretical study observed catalytic peptide dendrimers presenting shapes close to spheres [Javor 2009]. However, experimental studies on the same systems obtained measures for the compactness (using diffusion NMR spectroscopy) suggesting the existence of both molten globule and more conformational disordered states [Javor 2007].

For a more comprehensive survey on the subject of peptide dendrimers, the reader is referred to some excellent articles [Sadler 2002; Delort 2006; Maillard 2009; Sadler 2002; Niedrhafner 2005; Veprek 1999(a)].

[2] Multiple antigenic peptides (MAP's) are the most common examples of this class [Tam 1988].

1.8. Peptide Dendrimers Applications

Merging the peptide and dendrimer fields was expected to produce synergistic effects. Features of peptide dendrimers include the following:

1. A protein-like structure which can act as a receptor by adapting to the shape of natural ligands [Javor 2007; Johansson 2008];
2. A polyvalent structure that enables simultaneous interactions between two or more ligands and receptors of the same type resulting in amplification of function [Johansson 2007];
3. Biocompatibility, which can minimize their cytotoxicity [Crespo 2005];
4. Water solubility, which is crucial for macromolecules designed for systemic administration [Darbre 2006];
5. Increased resistance to proteolysis, caused by their high degree of branching [Bracci 2003; Sommer 2009];
6. Biodegradability, which can circumvent problems related with polymer degradation [Crespo 2005].

Nowadays peptide dendrimers play an important role in diverse areas including anticancer, antimicrobial and antiviral drug research; catalysis models, protein and cofactor binding, protein models, gene delivery, among others [Niedrhafner 2005].

In medical diagnostics, peptide dendrimers have been used as contrast agents for magnetic resonance imaging (MRI), magnetic resonance angiography (MRA), fluorogenic imaging, and serodiagnosis [Stiriba 2002]. In therapeutic applications, they have shown promise as vehicles for delivering drugs, DNA, peptides, or proteins [Crespo 2005; Sadler 2002]. More recently, peptide dendrimers have also been used to construct liposomes and biocompatible surfactants [Crespo 2005].

There is an increasing interest in the synthesis of peptide-based dendrimers owing to what is thought to be their structural similarity to globular proteins. Some results are already available demonstrating dendrimers ability to act as catalysts and binding macromolecules [Douat-Casassus 2004; Johansson 2008].

Several million years of evolution were required to achieve the highly compact and well defined folds encountered in the globular proteins of modern organisms [Crespo 2005]. This evolution involved the iterative mutation of amino acids at positions which ultimately led to globular-shaped proteins, whereby hydrophilic residues are concentrated at the protein surface and hydrophobic side-chains are densely packed in the interior [Lockman 2005]. The geometry of some peptide dendrimers results in spherical macromolecules with high surface density and tightly packed congregation of amino acids in their inner core. Thus, compactness, an evolutionary advantage in natural globular proteins, is artificially achieved in peptide dendrimers via the layering of sequences based on natural amino acids outward from a symmetrical core [Lockman 2005]. Furthermore, modern synthesis methods allow the creation of enormous combinatorial libraries testing simultaneously different amino acids in multiple dendrimer positions [Clouet 2004(a)], thus reducing hugely the time needed to find the best configuration of amino acids for each task. In this manner multiple and efficient protein model systems can be created [Javor 2007; Darbre 2006; Douat-Casassus 2004; Clouet 2004(b)].

Obtaining productive interactions between amino acids within the dendritic topology remains however a complex problem, not very different from the situation in a folded protein, where small amino acid modifications often strongly affect function without significantly disturbing structure [Javor 2007; Sommer 2008].

1.9. Motivation/Aim of this work

Natural proteins acquire their function by organized folding, which leads to the formation of active sites. In 2003 Reymond and coworkers reported a synthetic strategy to circumvent the protein folding problem [Esposito 2003] based on the interplay between amino acids within the dendritic structure. These specific archetype of peptide dendrimers were formed by alternating functional amino acids with branching diamino acids [Clouet 2004(b)].

The peptide dendrimers obtained in this way have been continuously and systematically studied, rendering protein models for different natural systems. Some

examples include catalytic peptide dendrimers with esterolytic or aldolytic activity [Douat-Casassus 2004; Delort 2006; Javor 2007], multivalent lectin binding dendrimers [Johansson 2007] and peptide dendrimer models for vitamin B_{12} transport proteins [Sommer 2008].

The last ones are of particular interest to the work in hand. Many proteins incorporate cofactors to complement their functionality. In the case of metal-containing cofactors such as porphyrins or cobalamins, the protein acts as a protecting shell around the cofactor, which leads to altered reactivity at the metal center. Peptide dendrimers are among the first reported synthetic macromolecules, mimicking the cobalt binding in B_{12} dependent enzymes or transport proteins [Sommer 2008].

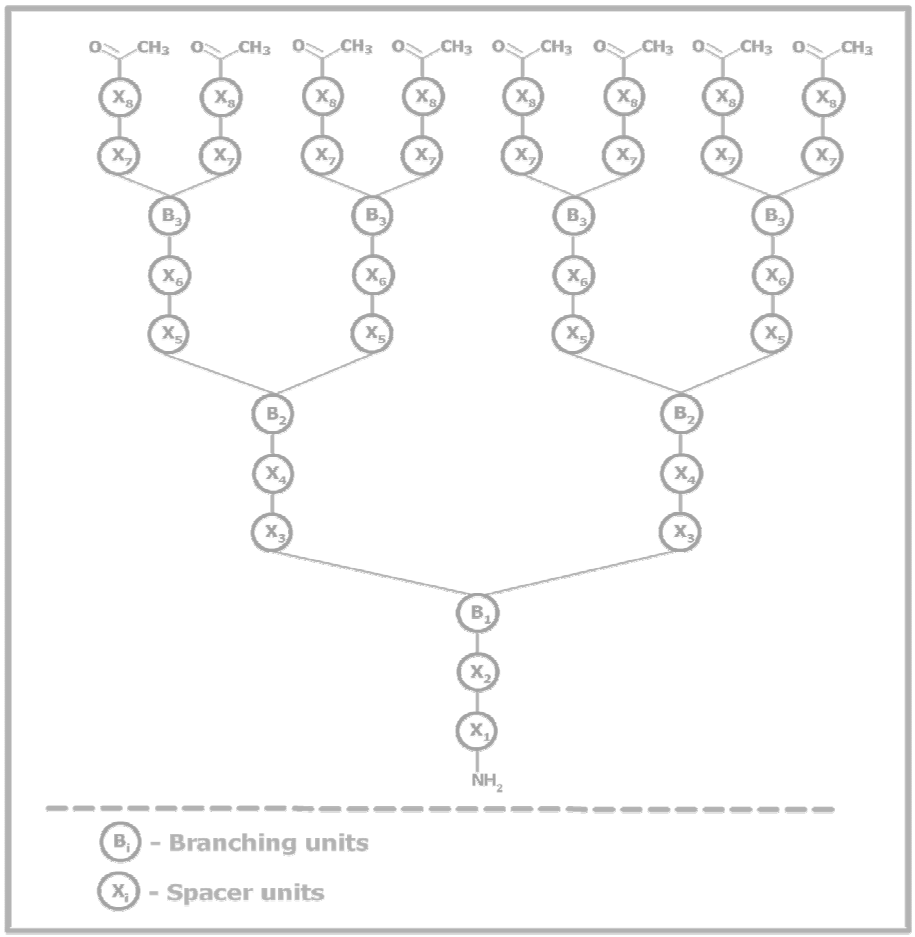


Figure 1.4 – Schematic representation of the Vitamin B12 transport protein dendrimer models architecture

Dendrimer schematics of the model for Vitamin B12 transport developed by Reymond et al. [Sommer 2008] which will also be used in the current thesis. It is a third generation architecture where each branching unit marks the beginning of a new generation. The represented dendrimers are built using $(X_{i+1}X_i)_2B_i$-type monomers, so that, each new generation of branching units doubles the number of peripheral groups. As can be observed in the scheme, all dendrimers are acetylated at the N-termini and aminated at the C-terminus.

These findings were the result of investigations carried out by Reymond and coworkers [Sommer 2008] in which they explored the molecular principles of cobalamin-peptidic ligand interactions using a dendritic architecture containing eight variable spacer units connected by three successive branching diamino acid units (see Figure 1.4) [Sommer 2008]. They developed a innovative combinatorial approach allowing the preparation of large libraries that were screened for B_{12} binding and catalytic activity[3] [Clouet 2004(a); Maillard 2009; Delort 2006; Johansson 2007].

The recognition and discrimination among a large number of amino acid sequences led to a specific peptide dendrimer model for a cobalamin-transport protein [Clouet 2004(a); Sommer 2008].

In their peptidic ligands, cofactor binding was mediated by a coordinating cysteine or histidine residue at the dendrimer core and was assisted by secondary interactions with the dendritic shells [Sommer 2008]. Specifically, this third generation peptide dendrimers consisted of proteinogenic amino acids arranged in a tree-like dendrimer topology by using a branching diamino acid (s-2,3-diaminopropionic acid (Dap) or lysine) at every third position in the peptide sequence. A schematic representation of these peptide dendrimer structures is presented in Figure 1.4. In their studies they were also able to develop a synthesis protocol for dendrimers up to the third generation (with good yield), identified and optimized the ideal number of spacers and the "best" variations in the incorporated amino acids (chosen from a predefined library. Therefore, not all the existing amino acids where tested/screened at every dendrimer position).

As a result from these multiple studies it is now known that dendrimers presenting the highest binding affinity to aquocobalamin, have some key positions that must be occupied by only a few types of amino acids in order to obtain the best functional chemical species.

For example, in case the cobalamin coordination residue present in the dendritic core is a cysteine (position X_2 in Figure 1.4) coupled with an aspartic acid (forming the core), then some of the outermost positions (positions X_8 and X_6 in Figure 1.4) must be, for the sake of binding optimization, occupied by glutamates.

[3] This approach intends to mimic natural evolution.

On the other hand, if the coordination to aquocobalamin is performed through histidine residues positioned in the first generation shell (position X_4 in Figure 1.4), then the glutamates must still be present at the second generation shell (position X_6 in Figure 1.4) in order to obtain a reasonable binding, but the outermost residues (position X_8 in Figure 1.4) have to be serines.

Dendrimers presenting the topology, constitution and coordination residues previously mentioned are interesting cases with potential biological applications, therefore they are the object of these thesis.

In this work we will study five chemically different peptide dendrimers (B1, B1H, B1HH, B1HHH and C1) containing either histidine, cysteine or both at aquocobalamin coordinating positions.

Information on the residues that compose each of our case studies can be found in Table 1.2. Figure 1.4 contains a scheme with the global structure of the dendrimers used in our study and in Appendix A we present their chemical structures.

Table 1.2 – Detailed residue composition for each dendrimer in study

This table presents detailed information on the dendrimers studied in the current work and should be interpreted taking into consideration the general dendritic architecture and residue positions presented in Figure 1.4. Discrimination between branching and spacer units is also evidenced. All dendrimers will be acetylated at the N-termini and aminated at the C-terminus (Figure 1.4). The residues and positions considered directly responsible for the binding to aquocobalamin are highlighted in gray.
$\mathbf{X_i}$(i=1,2,...,8) denotes a proteinogenic amino acid or 4-(aminomethyl)benzoic acid (**Amb**). $\mathbf{B_i}$ (i=1,2,3) represents the branching residues, s-2,3-diaminopropanoic acid (**Dap**) or Lysine.
The three letter code representing each proteinogenic amino acid corresponds to the standard nomenclature, but the corresponding names and one-letter codes can still be found in the Abbreviations table presented at the beginning of this thesis.

Dendrimer	Residues at each position										
	X_1	X_2	B_1	X_3	X_4	B_2	X_5	X_6	B_3	X_7	X_8
B1	Asp	Cys	Dap	Tyr	Amb	Lys	Ala	Glu	Dap	Ser	Glu
B1H	Asp	His	Dap	Tyr	Amb	Lys	Ala	Glu	Dap	Ser	Glu
C1	Ala	Arg	Dap	Thr	His	Dap	Tyr	Glu	Dap	Gly	Ser
B1HH	Asp	Cys	Dap	Tyr	His	Lys	Ala	Glu	Dap	Ser	Glu
B1HHH	Asp	His	Dap	Tyr	His	Lys	Ala	Glu	Dap	Ser	Glu

In the set of dendrimers used in this study, B1, B1H and C1 were actually synthesized by Reymond and coworkers [Sommer 2008] (the names used in the present work are actually the same name they used in theirs). Furthermore, in order to investigate the effect of additional metal-coordinating residues, we have also performed molecular simulations on two hypothetical dendrimers, B1HH and B1HHH.

We present four cases with the same general molecular architecture based on the B1 dendrimer [Sommer 2008], where B1 and B1H contain only one metal-coordination residue at the core, and B1H is obtained by replacing the coordinating cysteine by a histidine (position X_2 in Figure 1.4). This small change in a key residue leads to a significative decrease in the interaction with aquocobalamin (see Table 1.3). On the other hand, B1HH and B1HHH are obtained from B1 by replacing with histidines the two Amb residues (position X_4 in Figure 1.4) and, in B1HHH, also the cysteine core residue (position X_2 in Figure 1.4).

The binding affinities experimentally determined by Reymond and coworkers are presented in Table 1.3.

Table 1.3 - Experimentally determined aquocobalamin binding constant of some dendrimers

This table shows the available aquocobalamin binding constants for the dendrimers that will be studied in this work. The experimental determination of the values presented was carried out by Reymond and coworkers [Sommer 2008].
The residue sequence for each dendrimer is also presented, furthermore each color corresponds to a different generation.
Dap=s-2,3-diaminopropanoic acid; **Amb**=4-(aminomethyl)benzoic acid; **Ace**=Acetyl cap;
The three letter code representing each proteinogenic amino acid corresponds to the standard nomenclature, still the correspondent names and one letter codes can be found in the Abbreviations table presented in the beginning of this thesis.

Peptide Dendrimer	Sequence	Binding Constant $K_b/10^6$ [M^{-1}]
B1	$(AceGluSer)_8(DapGluAla)_4(LysAmbTyr)_2DapCysAsp$	5.0±0.800
B1H	$(AceGluSer)_8(DapGluAla)_4(LysAmbTyr)_2DapHisAsp$	0.083±0.011
B1HH	$(AceGluSer)_8(DapGluAla)_4(LysHisTyr)_2DapCysAsp$	n.a.[a]
B1HHH	$(AceGluSer)_8(DapGluAla)_4(LysHisTyr)_2DapHisAsp$	n.a.[a]
C1	$(AceSerGly)_8(DapGluTyr)_4(DapHisThr)_2DapArgAla$	0.022±0.002

[a]not available, theoretical model

Finally, we have studied a dendrimer that uses only s-2,3-diaminopropionic acid (Dap) as branching unit and contains two symmetric metal-coordinating histidines. This molecule presents a much lower binding affinity towards aquocobalamin than the former dendrimers (B1 and B1H) having a single a strategically positioned metal-coordination residue at the dendrimer core (see Table 1.3).

Despite the fact that this is a theoretical work and that we do not intend to formally study the binding of an aquocobalamin to the dendrimers in hand, the choice of the model dendrimers used was made based on the experimental binding affinities, so that we can not only search for conformational and folding preferences of the dendrimers, but also speculate on the issue of binding and on the consequence of different conformational characteristics and favored shapes on the results experimentally observed.

Accordingly we have selected some intriguing cases from the ones presented by Reymond and coworkers [Sommer 2008], where binding decrease was observed despite an increase of potential metal-coordinating groups. Moreover, we proposed two new cases combining additional coordinating residues (see Table 1.3).

An important consideration obtained by analyzing the binding constant values for the different dendrimers is that the peptide dendrimers mimic the binding mode of the vitamin B_{12} transporter without, however, rivaling with the protein in terms of cofactor-binding affinity (K_b(10^{10} M^{-1} for transcobalamin)[Sommer 2008].

1.10. Why use molecular dynamic simulations to study peptide dendrimers?

Having chosen the peptide dendrimers of interest it is now necessary to choose a way to study them. The conformational study of peptide dendrimer structures has lagged behind the fast progress in their synthesis and design. It is thought that the problem lies in the fact that these molecules possess an enormous number of energetically permissible conformations that rapidly interconvert in solution. As such, so far these structures have not yielded to direct structural determination by

crystallography or NMR [Javor 2009] affording large gaps in our understanding of these macromolecules. New insights into the rationalization and understanding of the experimental behavior of these molecules can be achieved with the help of theoretical approaches.

Determining the three-dimensional structure of dendrimer is a major challenge. Recently Müllen and coworkers have obtained the first single-crystal X-ray structure of a polyaryl dendrimer, a special type of particularly conformationally rigid dendrimer [Bauer 2002]. Most dendrimers, however, are conformationally quite flexible and have so far escaped a detailed structural/spatial characterization, some frustrating examples including the PAMAM, PPI and peptide dendrimers. To their structural characterization much time and effort has been devoted by different scientific groups throughout the world without obtaining substantial results. It is in this frustrating experimental scenario that molecular dynamics and computer simulation gain interest as a theoretical tool to understand the macroscopic behavior of dendrimers from microscopic interactions.

Like most macromolecules, dendrimers are not static structures [Javor 2009; Tanis 2009]. The impossibility of obtaining a x-ray structure provides by itself clues about the dynamic character of peptide dendrimers. The fact is that no such structure has been obtained by reason of difficulties in "seeing" certain parts of the molecule due to motion, disorder and flexibility.

Even if some experimental techniques can access certain degrees of macromolecular motion, molecular simulation itself is still the best tool to infer detailed atomic information on phenomenona that are not possible to scan directly with the currently available experimental techniques.

Since the 1970s computer simulation has been increasingly used by chemists, physicists, and molecular biologists to gain insight into molecular processes at a resolution often unreachable by experiment [Oostenbrink 2004].

Indeed, molecular dynamics simulations have been performed on various dendrimer types to understand their structural properties, with particular emphasis on PAMAM dendrimers [Lin 2005; Maiti 2004; Maiti 2005; Karatasos 2001; Gotze 2003; Han 2005].

Furthermore, theory and simulations may be used as direct guides in the course of the design and synthesis of chemical nanostructures such as dendrimers. Chemical structures can be mapped directly onto simulations and the results thereof may be used in turn to decide which structures should be made [Ballauff 2004].

Computer simulations are widely used in science, and dendrimers have been investigated using various theoretical models and computer simulations methods such as Monte Carlo (MC) and molecular dynamics (MD) simulations [Çagin 2000; Mansfield 2002; Lyulin 2000; Han 2005; Karatasos 2001; Maiti 2004; Klos 2009; Chen 1996; Freire 2008].

The main results of all these studies may be briefly summarized as follows: dendritic macromolecules can be successfully modelled with the current available computational resources [Ballauff 2004].

While coarse-grain MD studies on dendrimers [Murat 1996] were used to predict general properties, atomistic MD simulations on dendrimers were performed to provide 3D structures of specific dendrimers and to suggest possible interactions in dendrimer formation. Recently, more exhaustive atomistic MD simulations on dendrimers were carried out to study their solution behavior, potential use as catalyst supports, the effect of repeat unit flexibility, and the potential for several nanoscale applications [Çagin 2000; Ballauff 2004; Javor 2009; Gotze 2003].

These studies have helped to enlighten some of the main structure-property relationships and to ensure the applicability of standard simulation techniques to this peculiar field by studying model systems.

However, at least to our knowledge, with the exception of the study of Javor et al. [Javor 2009], there is still a lack of studies in what refers to peptide dendrimers with the sort of composition studied here, this is, containing naturally occurring amino acids and lysine or s-2,3-diaminopropionic acid as branching units.

In fact the first simulations of this sort of systems where carried out and published in the current year, alongside with the realization of our work [Javor 2009]. However, the sampling of conformational states employed in that study was somewhat crude, and was given little attention to the existence of structural trends. Moreover, the approach employed to validate the theoretical model is far from being consensual. Still, the publication of such article demonstrates the growing importance of the

peptide dendrimers field and the renewed interest in understanding the dynamic processes involving these structures.

With this in mind, the aim of this work is to explore the relation between the size and the structural properties of peptide dendrimer molecules and to proceed to a comparison with experimental findings. Special emphasis is placed on investigating the existence of conformational preferences representative of the examined dendrimers.

In here we recognize that the folding of macromolecules involves motion in a large range of length and time scales, and we define it not as a progressive pathway of unique single conformations, but rather as interconversions among ensembles of conformations in a back-and-forth progression from the unfolded to the folded state [Maisuradze 2009]. Thus, in the present work we characterize five third generation peptide dendrimers through long molecular dynamic simulations and analyze their conformational details and folding preferences, achieving a partial understanding of these captivating macromolecules.

Detailed information on what the reader needs to know about molecular dynamics (MD) methods and computational simulation are presented in the next chapter.

Chapter 2

Theory and Methods

In the first chapter we described the goals we set out to achieve along this thesis. Furthermore, we contextualized the scientific problem we propose to address (see section 1.9.). Taking the physical aspects of the problem into account we now describe the procedures undertaken to investigate it.

In the present chapter we will go through a few topics such as methodology and miscellaneous theoretical concepts of computer simulation, and, later on, the practical aspects and strategies employed in the current work.

This detailed description on the procedures and assumptions considered is supplied so that the reader can adopt an integrated perspective in line with the previous and next chapters.

Long molecular mechanics/molecular dynamics (MM/MD) simulations were used in this study to investigate structural properties of peptide dendrimers. A brief description of these concepts is thereby mandatory.

We hope that, with this, the reader can more easily understand the implications, developments and limitations of the presented work.

2.1. General Principles

2.1.1. Molecular Modeling: Description and Relevance

Computation based on models is playing an increasingly important role in biological sciences. Its relevance is supported by the fact that nowadays only a very limited number of methodologies addressing molecular systems are experimentally accessible. Crucial features of most molecules, such as dynamic behavior and inter/intra-molecular interactions at the molecular level, among others, are only approachable through computer simulations.

Computer simulation can therefore provide the framework for atomic scale studies of macromolecules and polymers, allowing the quantification of numerous properties and providing averages, distributions and time series of many definable quantities [Allen 1989].

Molecular modeling is the term applied to precise the variety of theoretical models that provide simplified descriptions of molecular systems mimicking some aspects of the behavior of molecules [Leach 2001].[4]

When applying this concept to macromolecular systems we are in fact describing complex systems composed of multiple particles in terms of a realistic atomic model. In here we are particularly interested in theoretical models that can be applied to systems composed of 10^4-10^5 particles. There are many different types of models and the choice of a particular one must be made according to criteria such as the

[4] This description, although quite embracing is still the subject of some dispute and consequently the contextual definition will depend on the author.

properties, the time scales of the events, and the number of degrees of freedom [Becker 2001].

In the current framework systems can be mainly described by two different approaches depending on the level of detail intended, *classical mechanics* and *quantum mechanics*.

In principle all biomolecular systems can be accurately described by quantum mechanics using the time-dependent Schrödinger equation, but in practice this approach is unfeasible for large molecules due to the complexity of the systems involved and computationally prohibitive calculations. Thus, approximations are necessary.

Using the simplifications provided by theoretical models, macroscopic properties can be understood and predicted[5] based solely on detailed microscopic (atomic scale) knowledge.

In this thesis we will apply a molecular mechanics theoretical model. This implies two major assumptions concerning the level of detail and the time span of the processes one is interested in.

The first assumption is that the abstract nature of electronic motion can not be effectively described, for large molecules, by the currently available methods, and therefore the Born-Oppenheimer approximation is invariably assumed to operate.

This enables the electronic and nuclear motions to be considered separately, and as the electrons present an endlessly smaller mass in relation to the nucleus, it is considered that the electronic part of an atom can rapidly adjust its position to any change in the nuclear positions. Consequently, the spatial trajectory described by a molecule can be solely considered as a function of its nuclear coordinates. When using this specific theoretical model an atom can be conceived as a single particle where the electrons are assumed to be on their ground state and their effect is incorporated in the nuclei [Leach 2001; Rapaport 2004].

---

[5] One of the factors that made a major contribution to the investment in computational chemistry is its ability to work as a predictive tool providing approaches for calculating several measurable and immeasurable properties of molecular systems.

The second assumption allows for systems that are of an intrinsically quantum nature to be dealt in a classical way. It states that the behavior (motion) of atoms (that are quantum mechanical entities) can be described with accuracy by simpler classical mechanics in opposition to the complex quantum mechanical treatment of motion (Schrödinger's equation).

Therefore, the atom nucleus moves according to the laws of classical mechanics [Allen 1989; Leach 2001].

Given the premises previously stated we will consider our macromolecules as "simple" classical mechanical systems composed of a group of atoms interacting with each other.

Addressing a molecular problem in such a way means that all atoms of the system are taken into account and the properties obtained from the molecular model can be related to the properties measured experimentally with the support of the theoretical principles of statistical mechanics [Leach 2001].

Brief descriptions of the essential concepts and equations inherent to the molecular modeling approaches employed in this thesis are presented next. Still, more detailed discussions on the subject can be found in books by Allen & Tildesley [Allen 1989] Leach [Leach 2001], Rapaport [Rapaport 2004], Berendsen [Berendsen 2007] and Becker [Becker 2001]. Good short reviews on this topic can also be found in references [van Gusteren 1990], [van Gusteren 2006] and [Karplus 2002].

2.1.2. Molecular Mechanics – Potential Energy Functions and Empirical Force Fields

A potential energy function is a mathematical equation that allows for the potential energy, $V(\mathbf{r}^N)$ (the potential energy of a system composed of N particles as a function of the atomic position vector $\mathbf{r}$ of each particle), of a chemical system to be calculated as a function of its three-dimensional structure [Becker 2001]. The forces of the individual particles of a system are deeply related to the gradient of this potential

energy function (see bellow), which is why these functions are commonly formulated as *force fields*.

This equation encompasses relatively simple terms that describe the physical interactions among all the particles in a system and dictates the structural features of dynamic molecules. Ideally, a force field incorporates all the relevant molecular interactions we believe will model the important degrees of freedom we are interested in [van Gusteren 1990; van Gusteren 2006; Ponder 2003].

Many of the potential energy functions in use today can be interpreted in terms of a relatively simple multiple-term depiction of the intra- and inter-molecular forces in the system [Leach 2001].

The most commonly used functions incorporate relatively simple equations describing each of the interactions contributing to the potential energy[6], a typical example being:

$$V(\boldsymbol{r}^N) = V_{Coulomb} + V_{vdW} + V_{bonds} + V_{angles} + V_{dihedrals} \qquad \text{Equation 2.1.}$$

$$V_{Coulomb} = \sum \frac{1}{4\pi\varepsilon_0} \frac{q_i q_j}{r_{ij}} \qquad \text{Equation 2.2.}$$

$$V_{vdW} = \sum \frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} \qquad \text{Equation 2.3.}$$

$$V_{bonds} = \sum_{bonds} \frac{1}{2} k_{ij}^{b} (r_{ij} - b_{ij}^{0})^2 \qquad \text{Equation 2.4.}$$

$$V_{angles} = \sum_{angles} \frac{1}{2} k_{ij}^{\theta} (\theta_{ijk} - \theta_{ijk}^{0})^2 \qquad \text{Equation 2.5.}$$

[6] In this definition we implicitly consider polarizability effects as absent. However, models accounting for these effects exist [Genzer 2004].

$$V_{dihedrals} = \sum_{dihedrals} \frac{1}{2} k^{\phi}_{ijkl} \cos(n_{ijkl}(\phi_{ijkl} - \phi^{0}_{ijkl})) \qquad \text{Equation 2.6.}$$

In here r_{ij} represents the norm of the vector $\mathbf{r}_{ij}$, defined between particle i and j as, $r_{ij} = \|\mathbf{r}_i - \mathbf{r}_j\|$. Equation 2.1. is one of the simplest potential energy functions that can reproduce the basic features of macromolecules (especially proteins and peptides) energy landscapes (the definition of energy landscapes will be introduced in the next chapter) with atomic detail. Changes in the degrees of freedom of a system can be viewed as movements along a multidimensional surface, whose height corresponds to the energy. In equilibrated systems, it is expectable that the most populated states are the low energy regions of the potential energy function (more information in provided in the next chapter). We will be particularly interested in *energy surfaces* later on.

It is the combination of a potential energy function (as Equation 2.1.) and all the parameters that go into it (k^{b}_{ij}, b^{0}_{ij}, k^{θ}_{ij}, θ^{0}_{ijk}, etc.) that constitutes a force field in the usual sense of the expression [Ponder 2003].

Equation 2.1. comprises two distinct types of atomic interaction among particles: *bonded* (depicted in blue) and *non-bonded* (depicted in green).

The non-bonded component is composed by the first two terms of Equation 2.1: $V_{Coulomb}$ (Equation 2.2.) and V_{vdW} (Equation 2.3.).

The $V_{Coulomb}$ term refers to the electrostatic interactions and uses a partial charge (q_i and q_j) placed on each atom so that interaction occurs via Coulomb's law. In many force fields, these interactions are considered only when atoms i and j are separated by more than three covalent bonds (see Figure 2.1). The V_{vdW} term depicts the van der Waals interactions which are represented according to a Lennard-Jones (12,6) potential reflecting the combination of dispersion and repulsion forces.

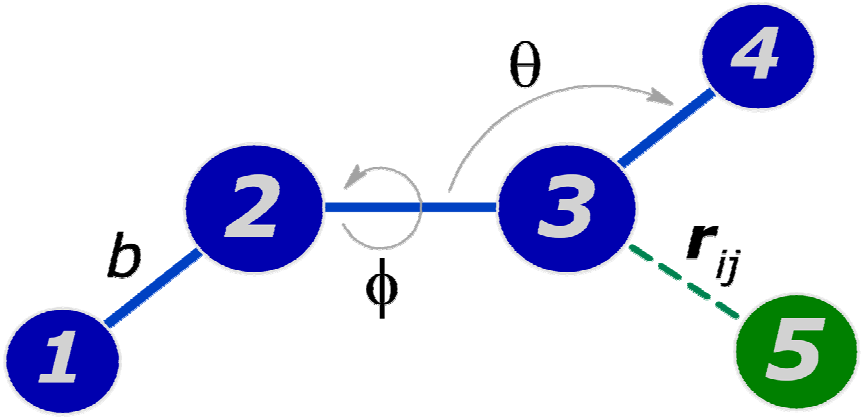


Figure 2.1 – Schematic view of the interactions involved in a potential energy function

Hypothetical molecules to illustrate the energetic terms included in Equation 2.1. Covalent bonds are indicated by blue solid lines and non-bonded interactions by a dashed green line.

The main molecule (Molecule X), depicted in blue, is composed by atoms 1 to 4. Another molecule (molecule Y), drawn in green, comprises atom 5.

Internal terms that occur in molecule X are: the bonds, b, between atom pairs 1-2, 2-3 and 3-4 (1,2 interactions); the bond angles, θ, involving atoms 1-2-3 and atoms 2-3-4 (1,3 interactions); and a dihedral or torsion angle, ϕ, described by atoms 1-2-3-4 (1,4 interactions). Molecule Y is involved in non-bonded interactions with all the atoms of molecule X characterized by different atomic distances (r_{ij}).

The last three terms of Equation 2.1. correspond to the bonded interactions (interactions between particles that are separated by no more then three chemical bonds) and are the ones related to covalent bonds. Hence we can distinguish three types of interactions characterizing these contacts: V_{bonds} (Equation 2.4.) where k^b_{ij} is the force constant and b^0_{ij} is the reference bond length; V_{angles} (Equation 2.5.) where k^θ_{ij} is the force constant and θ^0_{ijk} is the reference angle value; and $V_{dihedrals}$ (Equation 2.6.) where k^ϕ_{ijkl} is the force constant, n_{ijkl} is the multiplicity and ϕ^0_{ijkl} is the reference angle value. A diagrammatic representation of a hypothetical molecule in Figure 2.1 allows for visualisation of these different terms.

Only the bonded terms involve three- or four-body interactions, all non-bonded interactions are represented by pair-potentials, and all interactions are short-ranged except the electrostatic ones [Hess 2002].

Accordingly to what was previously stated, different parameters (k^b_{ij}, b^0_{ij}, k^θ_{ij}, θ^0_{ijk}, etc.) allow for different types of atoms and diverse molecular connectivities to be treated using the same set of equations (Equations 2.1.-2.6.). Thus the quality of the parameters employed to model each atom is questionable and optimizations keep showing up because, in fact, the parameters ability to reproduce experimental and quantum-mechanical data is what ultimately determines the accuracy of the results obtained.

Most of the force field parameters are taken from experimental data of small molecules and some others from theoretical calculations [van Gusteren 2006; Becker 2001; Ooostenbrink 2004; Hess 2006]. As such, *transferability* of the functional form and parameters is a trademark of the different force fields. In this context transferability means that the same set of parameters can be used to model a set of related molecules [Leach 2001].

Some considerations regarding force fields design and specificities are important and must be empathized:

(i). Force fields are empirical, which means that the set of parameters used to characterize each atom depend either on experimentally obtained data, quantum mechanical calculations, the force field developer, or in a conjunction of all three. But what mainly determines the design (and success) of a force field, is, from a pragmatic point of view, its ability to reproduce the experimentally observed properties of the systems. In fact, to a large extend even the form of the potential energy function is empirical, since it depends on considerations of what terms are and are not required to effectively model the system of interest [Becker 2001; Leach 2001].

(ii). When comparing or deciding which force field to use we must bear in mind that their quality can not be easily accessed and will depend on the type of system and properties one intend to investigate [Leach 2001].

(iii). The form for the potential energy function in Equations (1)-(6) represents a compromise between simplicity and the required chemical accuracy. However, more sophisticated force fields may present additional terms describing some interactions or molecular processes with further detail [van Gusteren 2006].

Force fields, force field parameterization and associated methodologies are by now well-enough established and there are many good textbooks covering the basics [Leach 2001; Becker 2001].

2.1.3. United-Atom Force Fields

Until now we have discussed force fields explicitly depicting all the atoms in a molecule (*all-atom force fields*), however there exist force fields that subsume some atoms (usually the hydrogen ones) into the atoms to which they are bonded. These are called the *united-atom force fields* [Leach 2001; Ponder 2003].

Normally, in united-atom models the hydrogens are treated as part of the non-hydrogen atom to which they are covalently bonded. As an example consider a methyl group: in this case it would be treated as a single atom, with the Lennard-Jones parameters and charges adjusted to account for the omission of the hydrogens, rather than as four individual atoms as would be the case in an all-atom model.

Although this approach could be applied to all hydrogens it is typically used only for the non-polar (aliphatic and aromatic) ones, because polar hydrogens take a key role in hydrogen-bond interactions.

Calculating all the interactions in a system is obviously a very computationally time-consuming process as, for instance, the number of non-bonded interactions scales with the square of the number of interaction sites. Therefore, by using united-atom force fields significant computational savings are possible without a major loss of detail [Leach 2001; Becker 2001].

2.1.4. Molecular Dynamics Simulation

2.1.4.1. Theoretical Basis

Macromolecules, in particular biomolecules such as proteins or lipids, are not static structures, and as a consequence most properties of molecular systems cannot be derived from static configurations of a system [Allen 1989].

The folding of a molecule for instance can only be ascertained if we have access to the myriad of states (ensemble of configurations) the system visits along a significant amount of time and with the theoretical support provided by *statistical mechanics*[7].

The laws of statistical mechanics govern the connection between the microscopic behavior and the macroscopic properties of molecular systems, providing the fundamental theoretical framework for molecular dynamics simulations.

An important consideration provided by statistical mechanics to the current work is that, in order to derive a molecular property, A, from a system we must have a collection of different configurations (*ensemble*) of that system [Chandler 1987]. The ensemble must be obtained under certain physical conditions (such as temperature, volume, number of particles, among others) from which we can average a macroscopic property, $\langle A \rangle_{ens}$, that can be compared against its counterpart experimental value.

In other words, using molecular simulations there are a number of properties that can be directly obtained across the simulation time, such as the velocities or positions of the atoms. However, this information cannot be compared with experimental data, because none of the available experimental techniques can provide such details. In fact, a typical experiment determines an average property, averaged over all particles

[7] A detail discussion on the issue of statistical mechanics is out of the scope of this thesis. For further information the reader is pointed to an excellent book written by David Chandler [Chandler 1987].

in the system, and during the time taken to perform the measurement (meaning that is also an average over time). Moreover, the measure itself changes the biomolecule, in a way that can not be predicted (Heisenberg's uncertainty principle).

As such, we must aim to compute not all the states of the biomolecule, but rather a representative set of equilibrated conformations, reflecting the overall relative populations of the states accessible to the system (an *ensemble*). To collect an ensemble, a set of configurations is needed and this entails employing a method for sampling an extremely large conformational space. This can be done using MM/MD and the framework of statistical mechanics.

The basic assumption underlying statistical mechanics is the *ergodic hypothesis*. This hypothesis states that following a single molecule across time (for a sufficiently large time interval, and in an equilibrated system), is identical (in terms of sampling the system states, or for our purposes the conformations) to capturing the individual features of a very large ensemble of identical molecules, at a single instant. The major implication of the ergodic hypothesis is that: the averages taken along time will (for a sufficiently large time interval) reflect the same trends as the average over an ensemble. Implying that the ensemble average, for a certain property, A, is independent of time. Accordingly we can calculate the average value of A, by considering "simply" the entire set of A values of each conformation in the ensemble at a single instant of time [Allen 1989; Leach 2001; Chandler 1987]. Furthermore, we can write the average of A as:

$$\langle A \rangle_{ens} = \frac{1}{Z} \sum_{z} A(z), \qquad \text{Equation 2.7.}$$

where z represents each molecule in the ensemble, at the instant of time considered, and Z is the total number of molecules.

Furthermore, as the ensemble accounts for every potential state the system can visit (from now on represented as s), and contains the relative number of times the system "visits" each state (meaning that the states most "represented" in the ensemble have a larger probability of occurring), we can define $n(s)$ as the number of times each state is "visited", and write the result of a certain property (in here A) as:

$$\langle A \rangle_{ens} = \frac{1}{Z}\sum_{z} A(z) = \sum_{s} n(s)A(s) = \frac{1}{Z}\sum_{s} P(s)A(s), \qquad \text{Equation 2.8.}$$

where $A(s)$ is the value of A for a state s, and $P(s)=n(s)/Z$ is the probability that a molecule in the ensemble is in that state. As a consequence, we observe that all that is needed to adequately describe the system, is a way to determine the relative probability of each state. This is where MM/MD acts. MM/MD generates the set of states (or conformations) needed to define the ensemble. The ensemble depends, of course, on the conditions chosen to simulate the system (pressure, temperature, among others).

In particular, for the ensemble we have used to simulate peptide dendrimers (the isothermal-isobaric or NPT ensemble; see further in the text), the probability of finding the molecules in a state s, is given by [Chandler 1987]:

$$P(s) = \frac{e^{-(E_s + pV_s)/RT}}{\Delta(p,T)}, \qquad \text{Equation 2.9.}$$

where,

$$\Delta(p,T) = \sum_{s} e^{-(E_s + pV_s)/RT}, \qquad \text{Equation 2.10.}$$

and V_s, E_s, p, R, T, represent the volume of state s, the energy of state s, the pressure, the gas constant and the temperature respectively.

This sophisticated technique has been used for some decades now with proven results [Becker 2001; Karplus 2002; van Gusteren 1998]. It makes the simulation of dynamic systems feasible through efficient computational calculations and allows for the gathering of a large collection of temporally related configurations.

Molecular dynamics solves the classical equations of motion for a system composed of N atoms, which results in a trajectory describing the atomic positions of all the particles in the system throughout the simulation time [Allen 1989].

From these atomic trajectories a variety of properties can be calculated [Allen 1989].

Accordingly, in MD simulations it is Newton's classical equation of motion that is solved. In its most simple form, Newton's equation of motion (or Newton's second law) states that for a system of N interacting atoms there exists a relationship between the atoms mass (m), the acceleration (**a**) and an applied force **F** according to

$$\mathbf{F}_i = m_i \cdot \mathbf{a}_i, \quad i = 1,2,...,N. \qquad \text{Equation 2.11.}$$

This can also be written in terms of atomic velocities, $\mathbf{v}_i$ (Equation 2.13.), and atomic position vectors, $\mathbf{r}_i$ (Equation 2.12.) with respect to time as:

$$\frac{\partial \mathbf{v}_i}{\partial t} = \frac{\mathbf{F}_i}{m_i}, \quad i = 1,2,...,N, \qquad \text{(Equation 2.12.)}$$

$$\frac{\partial^2 \mathbf{r}_i}{\partial t^2} = \frac{\mathbf{F}_i}{m_i}, \quad i = 1,2,...,N, \qquad \text{(Equation 2.13)}$$

The force, $\mathbf{F}_i$, on each particle in the system will also be determined by the gradient of the potential energy, V, calculated relative to the position of that atom, which is a function over all the atomic coordinates of the system ($\mathbf{r}_1, \mathbf{r}_2, ...\mathbf{r}_N$)

$$\mathbf{F}_i = -\nabla_{\mathbf{r}_i} V(\mathbf{r}^N) = -\frac{\partial V}{\partial \mathbf{r}_i}. \qquad \text{Equation 2.14.}$$

Of relevance is the fact that forces are vectorial quantities while the potential energy, V, is a scalar quantity.

Therefore, in a MD algorithm, the forces are calculated as the negative derivatives of all the analytic expressions describing the potential energy function [Allen 1989; Leach 2001; Becker 2001]. Once the force acting on all atoms is calculated we can integrate Newton's equation of motion and obtain the particle's positions and velocities using numerical methods.

We can briefly describe MD as a method that allows for the equations of motion to be numerically solved and as a result provides information on the time dependence and magnitude of fluctuations in both positions and velocities for a multiple particle system [Allen 1989; Leach 2001; Berendsen 2007; Becker 2001]. It is a deterministic

method by which we mean that, the state of the system at any future time can be predicted from its current state.

A schematic overview of a general simulation algorithm is presented in Figure 2.2.

Initial Conditions – Characterize the starting conformation	
1	Set an initial configuration with positions $\mathbf{r}_i$ for all the atoms in the system
2	Choose an initial set of velocities $\mathbf{v}_i$ for all the atoms in the system
3	Define a potential energy function V according to all atomic positions
Compute Forces – The forces are combined with the current atomic positions and velocities to generate new positions and velocities for the next step	
4	The forces acting on the atoms are computed from the potential energy expression (using a force field) $\mathbf{F}_i = -\nabla_{\mathbf{r}_i} V(\mathbf{r}^N) = -\frac{\partial V}{\partial \mathbf{r}_i}$, Equation 2.14. calculating the forces between non-bonded atoms plus the bonded interactions contributions $\mathbf{F}_i = \sum_j \mathbf{F}_{ij}$ Equation 2.15.
5	The potential and kinetic energies are determined
Update the systems configuration – move the atoms to new positions	
6	The forces calculated are assumed to be constant during a short time interval, called the time step Δt. Newton's equation of motion are solved and motion is simulated through computation of new positions for the atoms a short time ahead (Δt) $\frac{\partial^2 \mathbf{r}_i}{\partial t^2} = \frac{\mathbf{F}_i}{m_i}$ Equation 2.16.
7	Compute new velocities and accelerations for the atoms
If required repeat the steps 4 to 7 using the new positions to update the computed forces. This procedure must be iteratively executed until you collect enough data (total simulation time).	

Figure 2.2 – Step-wise scheme of a MD algorithm

Molecular dynamics simulations and an appropriate force field is a powerful combination that can provide considerable contributions to the existing scientific knowledge by:

1. Help in the construction of a synergistic understanding and interpretation of experimental results;
2. Estimation and prediction of physical properties experimentally inaccessible;
3. Investigate the conformational properties of flexible molecules;
4. Insights into the dynamic fluctuations of biomolecular processes.

The problems and limitations of MD simulations and methodologies are thoroughly discussed elsewhere [Leach 2001; Allen 1989; Rapaport 2004; Becker 2001; van Gusteren 1990; van Gusteren 2006; Berendsen 2007].

Other theoretical concepts will be introduced in the next chapter as we go through the analysis of our results and they become necessary.

2.1.4.2. Practical considerations

Despite a well defined support, MD requires some user supplied input parameters to solve the equations of motion, typically a set of initial coordinates $\mathbf{r}_0$ and a set of initial velocities, $\mathbf{v}_0$.

These parameters are necessary because only by knowing the position (**r**), velocity (**v**) and acceleration (**a**) at a time t, we can numerically obtain these same quantities at time $t+\Delta t$.

Usually in MD the initial coordinates of a system are obtained from experimentally determined molecular structures (mainly from crystallographic experiments), but alternatively, if no experimental structure is available, the initial coordinates can be generated using computer models through a variety of modeling techniques. The only relevant information available regarding the initial velocities is the system temperature (*T*), which controls the velocity distribution. In the absence of

information the initial velocities are normally randomly assigned [Leach 2001; van der Spoel 2005(b); Becker 2001].

Most molecular simulations can be split into five distinct stages in what we might call a *simulation protocol*. Each stage is described bellow:

1. Preparation: Establish an initial conformation for the system, perform energy minimizations if necessary.
2. Initiation: Assign initial velocities and gradually scale them to the desired temperature, accompanied by restrained dynamics.
3. Equilibration: A dynamic simulation run that lasts until the system achieves stability and stops demonstrating erratic fluctuations. The equilibration of a system can be evaluated using some key thermodynamic and structural properties of the system. For inhomogeneous systems this stage may take many time steps.
4. Production: When the system is at equilibrium, the dynamic simulation is considered reliable and the production phase commences. During this stage simple properties are calculated and the system's configuration is determined and stored at regular time intervals. Normally this is the longest stage, taking from several hundred picoseconds up to hundreds of nanoseconds.
5. Analysis: The trajectories and configurations are examined to calculate other properties and obtain further information about the system. By averaging over the equilibrated trajectory many macroscopic properties can be determined.

When designing a simulation protocol for a practical case, such as the work presented in this thesis, one needs to understand the practical limitations underlying molecular dynamics simulations and the conclusions that can be withdrawn from these studies. In fact, with the currently available computational power not all molecular systems can be simulated with atomic detail or for the desired amount of time. A compromise between a reasonable time period and an acceptable description of the system (in terms of particle number) must be reached.

The total time span for MD studies nowadays ranges from hundreds of picoseconds up to hundreds of nanoseconds[8].

The systems size that can currently be considered range up to 10^5 or 10^6 atoms [Weber 2000; van Gusteren 2006], which is still very small when compared to Avogadro's number, that is, macroscopic sizes. As a typical molecular simulation is carried out using a single macromolecule surrounded by solvent, there as to exist a boundary at the edge of the "simulation box" (system). This idea is the opposite of the behavior exhibit by real systems that are basically infinite with respect to the number of particles that define an observable or experimental "sample". In MM/MD studies this issue gains importance as the modeling of boundaries and surface effects has a large impact on the calculated properties [Weber 2000; van Gusteren 2006].

The classical way to minimize this "edge effects" in a finite system is to apply *periodic boundary conditions*. This means that the atoms of the system are put into a space-filing box that is surrounded by an infinite number of translated copies of itself (see Figure 2.3 for a schematic description). Thus there are no boundaries to the system. Although this is the best approximation to simulate an infinite system, the use of periodic images removes the boundary effects while introducing a high degree of periodicity [Leach 2001].

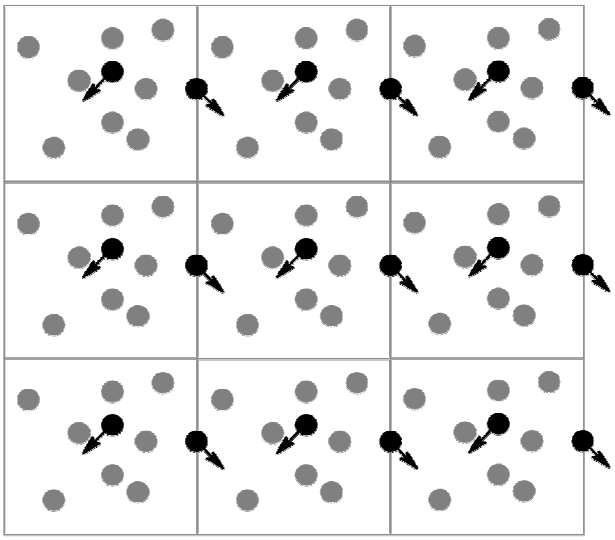


Figure 2.3 – Periodic boundary conditions in two dimensions.

A two-dimensional periodic system. Molecules can enter and leave each box across each of the four edges. In a three-dimensional example, molecules would be free to cross any of the six cube faces.

[8] Although some scientific groups have already reported microsecond simulations using enormous computational clusters [Kepleis 2009].

2.2. Overall features of the simulations performed

As previously mentioned (see section 1.9.), herein we present long MM/MD simulations of five topologically different dendrimers each in a box solvated with water. Further details are presented in Table 2.1.

Table 2.1 – Composition of each dendrimer in study

This table presents the amino acid residues that compose each of the dendrimers that will be studied in the current work and should be interpreted taking into consideration the general dendritic architecture and residue positions presented in Figure 1.4 and Table 1.2. Discrimination between generations is evidenced through the use of different colors. All dendrimers will be acetylated at the N-termini and aminated at the C-terminus (Figure 1.4).

The three letter code representing each proteinogenic amino acid corresponds to the standard nomenclature, whose correspondence can be found in the Abbreviations table presented in the beginning of this thesis.

Peptide Dendrimer	Sequence
B1	$(AceGluSer)_8(DapGluAla)_4(LysAmbTyr)_2DapCysAspNH_2$
B1H	$(AceGluSer)_8(DapGluAla)_4(LysAmbTyr)_2DapHisAspNH_2$
B1HH	$(AceGluSer)_8(DapGluAla)_4(LysHisTyr)_2DapCysAspNH_2$
B1HHH	$(AceGluSer)_8(DapGluAla)_4(LysHisTyr)_2DapHisAspNH_2$
C1	$(AceSerGly)_8(DapGluTyr)_4(DapHisThr)_2DapArgAlaNH_2$

To achieve this goal we are going to need specific software applications, computational power and a detail protocol that ensures an accurate conformational sampling.

As a remark, it must be emphasized that all dendrimers were simulated using the sequence with which they were (or would be in case of B1HH and B1HHH) synthesized, namely using a neutral amine group attached to the core as an end cap

(C-terminus amine), and acetyl (Ace) molecules connected to the shell end residues and acting as protecting caps (N-termini acetylated) [Sommer 2008].

A detailed overview of each stage of the setup and simulation procedures is presented over the next topics. The chapter ends with a scientific article-like outline of the methodologies used.

2.2.1. Computational Resources

The calculation resources employed were provided by the computational cluster available at Instituto de Tecnologia Química e Biológica (ITQB) [Molecular Simulation, ITQB; Protein Modeling, ITQB].

All simulations presented here were performed using the GROMACS 4.0.2 package [Berendsen 1995; Lindahl 2001] which is one of the most used and probably fastest [Hess 2008; van der Spoel 2005(a)] engines to perform molecular dynamics simulations and energy minimizations of biomolecules. Still, the total computational time employed was approximately 50 000 hours using six CPUs for one year.

2.2.2. Integration of equations of motion

The equations of motion were numerically integrated with the *leap-frog algorithm* [Leach 2001; van der Spoel 2005(b)] using a time step (Δt) of 2 fs, the neighbor lists employed were updated every 10 fs and the data regarding the system was saved at every 10 ps.

2.2.3. Boundaries

In all the simulations performed periodicity was considered through the use of periodic boundary conditions in all the dimensions of the Cartesian space.

Furthermore, in the case of our simulations we are particularly interested in the behavior of the central solute molecule (a dendrimer from the set previously described – Table 2.1) and as such we want to minimize the number of solvent molecules and associated calculations. Therefore we have decided to use a rhombic dodecahedron unit cell. This space-filling shape is very similar to the one of a sphere and allows significative savings in computational time (see below in Figure 2.4, step 5)[van der Spoel 2005(b); Becker 2001].

2.2.4. Long-Range Interactions

There are several methods for treating electrostatic forces in MD simulations with varying degrees of physical fidelity and computational efficiency. Herein we chose to use the *Reaction-Field* (RF) approach to model long-range electrostatics. Any method that does not take long-range electrostatics into account faces several artifacts [van der Spoel 2006; Yonetani 2005].

With this approach the electrostatic interactions between particles below a certain cut-off (we have selected a cut-off of 1.4 Å) are calculated explicitly and interactions above the cutoff are modeled as a dielectric continuum (in here we use a dielectric constant of 54.0 [Smith 1994]). This allows for significative computational savings without qualitative loss of information. In fact, the RF approach has been shown to reproduce satisfactorily the results of the other electrostatic methods [Machuqueiro 2006].

Both the van der Waals and the electrostatic interactions were modeled using a *Twin-Range Cutoff*. This means that the interactions among two particles whose

distnace is between the lower and higher cut-offs are calculated and it is assumed they remain constant for a given amount of time, after which the value for the interaction is recalculated and a list of interacting particles is updated [Leach 2001]. The interaction of particles whose distance is below the lower cut-off are computed at all time steps. We have considered here a cutoff between 8 and 14 Å and updates every 10 fs.

2.2.5. NPT Ensemble

Our simulations were performed in the isothermal-isobaric (NPT) ensemble, meaning that the number of particles, the pressure and the temperature remain constant in the course of the simulation [Becker 2001; Chandler 1987].

This line of action tries to model the experimental conditions in which dendrimers are obtained.

To simulate these conditions with GROMACS we need to employ a thermostat and a barostat. There are several methods available for temperature coupling; in here both solvent and solute were separately coupled to external temperature baths at 298.15 K through the use of a Berendsen coupling with a relaxation time of 0.1 ps. The *Berendsen temperature bath* acts as a thermic energy source adding or removing heat from the system as needed [van der Spoel 2005(b); Berendsen 1984].

We have also decided to model pressure using this approach; therefore we have used a Berendsen isotropic pressure coupling at 1 atm, with a relaxation time of 0.5 ps and a isothermal compressibility of 4.5×10^{-5} bar^{-1}.

2.2.6. Constraints and Restraints

Simulations of macromolecules with atomic detail, including explicit solvent in a periodic unit-cell, are limited in length by the computational power available and in

that context every simulation protocol represents a compromise between high accuracy and feasibility.

In the case of MM/MD the integration steps (or time step) must be short due to the high motion frequencies of the system. A way to make MM/MD simulations more quick and robust is to constrain certain degrees of motion associated with high motion frequencies so that one can use a higher integration step without jeopardizing the accuracy of the results [Leach 2001; Becker 2001].

Consequently, the entire set of MM/MD runs presented here where performed with all covalent bonds constrained using the LINCS (Linear Constrain Solver) algorithm [Hess 1997] for the atoms of dendrimers, and the SETTLE algorithm [Miyamoto 1992] to constrain the solvent (water) molecules.

It should be noted that, in the context of molecular simulation *constraints* and *restraints* are two completely different concepts. Constraints deal with a requirement the system is forced to satisfy (for example a bond length that is kept constant throughout the entire simulation) while restraints concern an "encouragement" to adopt a particular value (for example, a bond is able to deviate from its reference value but it faces a energy penalty for this) [Leach 2001].

In the present work we use them both (constraints and restraints) but while constraints are ubiquitously used on covalent bonds (bond-constraints) the restraints (position-restraints) will only be used in some initiation and minimization procedures (and not always applied to all the atoms). When necessary we will explicitly depict if restraints are being used and whether they are being applied under all atoms or just to the alpha carbon ones.

The position restraints employed here (always using a force constant of 1000 $kJ.mol^{-1}.nm^{-1}$) restrain the selected atoms to their reference positions during energy minimization and initiation, thus avoiding drastic rearrangements in the dendritic structure due to the presence of a non-equilibrated solvent.

2.2.7. Sampling

When performing an MM/MD study a recurrent question concerns the sampling procedure: how do we ensure that all the relevant states of a macromolecule are being correctly sampled during a run?

There is no easy answer for this question and in fact only the analysis of the results can address it. Still some procedures (heuristic or systematic) have been described to increase the probability of successfully sample a macromolecule's conformational space [van Gusteren 1998; Karplus 2002].

Conformational sampling is a process used to generate a collection of molecular conformations that will later be analyzed [Becker 2001]. Ideally, all stable conformations of a molecule should be accounted for, however macromolecules are far too complex making this enumeration impractical [Becker 2001].

In this work, our goal is to use MM/MD simulations to map the conformational behavior of peptide dendrimers. Therefore, we need to develop a protocol for generating an ensemble (a representative set of conformations; see section 2.1.4.1) that is statistically robust and thus allows the calculation of meaningful averages of some characteristic properties of the system in hand (e.g., radius of gyration, solvent accessible surface, radial distribution functions, among others).

In our particular situation, we simulated dendritic molecules and consequently some other questions arose. Although branched macromolecules have been simulated through MM/MD several times before (see section 1.10) the fact is that dendrimers, with the topological characteristics we intended to study, have been simulated only once [Javor 2009] and using a very different approach to search its conformational space. As a result, three main questions regarding the conformational sampling protocol need some discussion:

1. There is no experimentally determined structure of the system, so which structure should we use? Does the system loose memory of its initial conformation in the course of the simulation?
2. What are the time scales for relevant transitions to occur? Are they inferior to

the simulation time?

3. The process of macromolecular folding involves many degrees of freedom and is governed by small energy fluctuations on the order of a few multiples of k_BT (k_B=Boltzmann's constant) [Daura 1999]. Is the accuracy of the existent atomic interaction functions capable of discriminating these small differences?

To address these questions some considerations are necessary. First we must emphasize that the folding of macromolecules is best understood from the vantage point of its underlying *energy landscapes* (or *hypersurfaces*)[Frauenfelder 1998], which are surfaces defined over all conformational space, indicating the potential energy of each and every possible conformation[9] of the molecule [Becker 2001].

For complex molecules involving several degrees of freedom this potential energy landscape typically contains multiple energetic minima. And if we are not careful the simulation of the system may start trapped in one of such minima.

To ensure a (as good as possible) protocol that generates a statistically robust ensemble, we have first employed an energy minimization (procedure to locate a local minimum) to the modeled initial configuration of each peptide dendrimer, in this way stabilizing the systems (dendrimers plus solvent) and enabling the MM/MD simulations to start.

Energy minimization is widely used prior to molecular simulations to prepare systems for different calculations, being specially recommended for simulations of complex systems such as macromolecules or large molecular assemblies [Leach 2001].

After that we performed an iterative initiation protocol heating and cooling the system forcing it to overcome possible energy barriers. Furthermore, we have altered (for a brief period of time) the atomic electric charges of the dendrimer's atoms from their reference values to a positive and common one (+0.1 e) thereby promoting the repulsion among all the atoms and getting a generic structure without physiological

[9] By a *molecular conformation* we mean a particular arrangement of the atoms comprising the system of interest, typically described using Cartesian coordinates [Becker 2001].

meaning but that is the most "stretched" conformation for each dendrimer (we have affectionately named it *the hedgehog structure*).

In principal this "stretched" structure does not represent a conformational minimum and none of the intra-molecular interactions imposed by the starting conformation would be present. This is, in our opinion, the best chance to avoid that the conformation search occurring through MM/MD, starts trapped in an energetic minimum. Moreover, given the experimental evidences available for these dendrimers, this is the less biased approach because it allows the simulations to depart from states that are, in principle, the least favored ones.

In addition to the approaches previously described, we have also assigned random initial velocities (at the beginning of the initiation stage) to ten different simulations for each dendrimer, thus effectively producing different replicates.

This methodology to perturb a system is very common (e.g., [Oliveira 2005]) and induces the system into a direction in conformation space that strongly depends on the velocities generated[10]. Given the fact that all replicates are performed with identical conditions but the systems are driven into different directions, the trajectory-specific features resulting from the initial choice should be eliminated by replicate averaging, thus validating the conformational sampling approaches used (i.e., systems that depart from different directions cover similar areas of the conformational space).

Based on these considerations we have decided to use ten MM/MD replicates, each 100 ns long, for each of the five dendrimers under study (see Table 2.1).

[10] In fact, by performing ten simulations of the same initial 3D structure but employing different initial velocities we are implementing the use of replicates, as is commonly done in experimental protocols. The justification for this option is thoroughly discussed in a article by Soares and coworkers [Oliveira 2005] but can be summarized as: "a combined approach between temporal and ensemble (replica) averaging helps to mitigate the sampling problem".

2.2.8. GROMOS96 53A6 Force Field

A major issue when performing molecular simulations is the choice of a force field. Currently, a wide variety of force fields are available and therefore one must carefully choose the most appropriate for the particular task at hand. Most important in this selection process is knowledge of the information to be withdrawn from the computational study.

In here, we intended to study the conformational variability presented by different examples of peptide dendrimers and, for that reason we aimed at a significant level of atomic detail. Still, it is essential that the computational time needed to accomplish this task is as small as possible. Therefore the force field chosen must reflect a compromise between accuracy, proficiency and time economy.

Hence we have decided to perform our simulations using the GROMOS96 53A6 [Oostenbrink 2004] united-atom force field to model the peptide dendrimer particles and explicitly modeled the solvent molecules through the use of the SPC water model [Berendsen 1981].

The GROMOS96 53A6 is a force field designed and optimized for biomolecular simulations that has been extensively used and tested in simulations of a wide variety of proteins, nucleotides, sugars and lipids [Ooostenbrink 2004; Oostenbrink 2005].

The adequacy of the GROMOS96 53A6 force field to deal with amino acid residues is well documented an its validation has been the topic of multiple studies with its strengths and weaknesses being properly described [Oostenbrink 2005].

At the time we started our simulations, this force field had not been used to simulated peptide dendrimers with the particularities presented here. However, by using this force field Reymond and coworkers [Javor 2009] have successfully reproduced experimental hydrodynamic radii for single-site esterase peptide dendrimers.

2.2.9. Water Model

When performing a simulation, different levels of interactions must be considered: solute-solute interactions (peptide dendrimers intra-molecular contacts), solute-solvent interactions (water-dendrimer contacts) and finally solvent-solvent interactions (water-water). As can be observed through Table 2.2 (see below) all the simulated systems are composed by one solute molecule (a peptide dendrimer) and numerous solvent molecules (waters), as such the most computationally expensive process lies in the calculations associated with the water-water interactions [Leach 2001]. To minimize the computational cost related with this interactions we have employed a commonly used water model in biomolecular simulations, the SPC model [Berendsen 1981; Hess 2006].

SPC is a three-point charge model (at the hydrogen and oxygen positions) with a Lennard-Jones 6-12 potential on the oxygen atoms only and optimized for use with a reaction-field treatment [Berendsen 1981; van der Spoel 1998(a)].

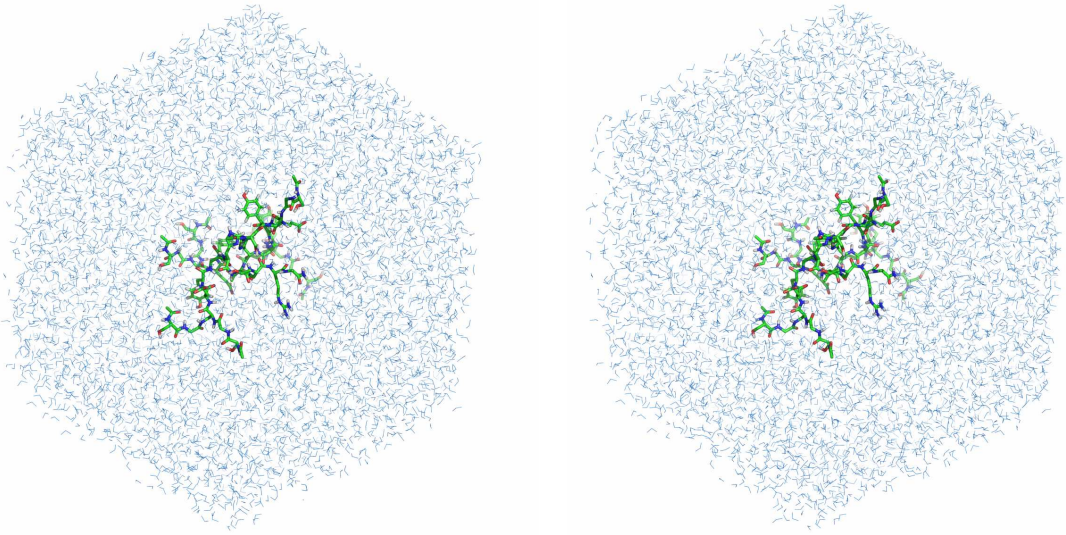
This model presents a simple but effective way to explicitly depict the behavior of water in our system.

2.2.10. Layout of the MD simulations

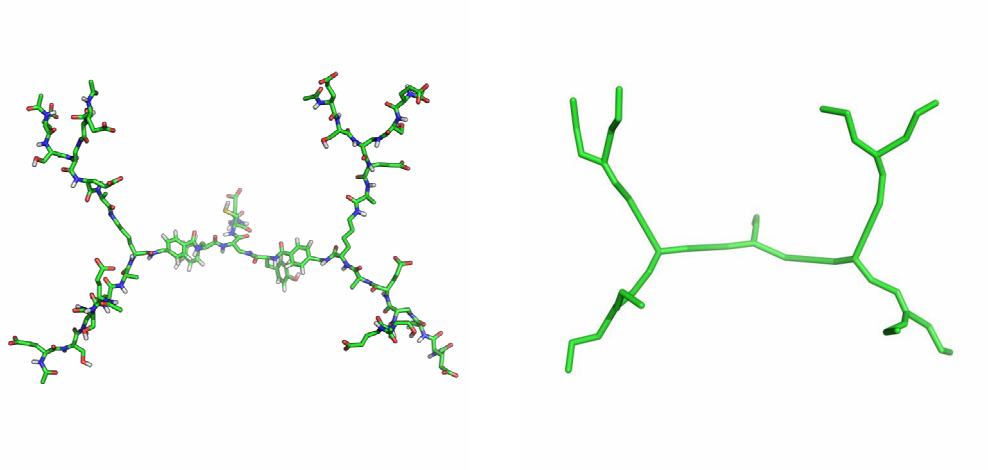
The Figure 2.4, presented below, shows a resumed and chronologically ordered description of all the steps preformed to efficiently design and obtain the molecular dynamics trajectories that will later be analyzed (see Chapter 3).

Step	Description
	Simulation Stage: Generation of input files and parameters
1	Construct a viable three-dimensional atomic structure for each peptide dendrimer. **B1 dendrimer**  **B1H dendrimer**  **B1HH dendrimer**  **B1HHH dendrimer**  **C1 dendrimer**  **Legend:** 3D Structures of the initial conformations built with PyMOL [DeLano 2002]. The pictures on the left represent atoms as spheres and colored accordingly - H(gray), O(red), N(blue), C(green), S(yellow). The aromatic and aliphatic hydrogens are not represented. The pictures on the right represent a "skeleton-structure" of the peptide dendrimers, i.e., a line is drawn between each consecutive alpha carbon atoms. This was the representation chosen because its simplicity removes the "clutter" created by depicting all the atoms. Also present is the line connecting the nitrogen atom of the amino cap and the subsequent alpha carbon.

Continues on next page

2	Choose a Force Field suitable for the simulation of peptide dendrimers.
3	Parameterize the branching and spacer residues attending to the force field previously chosen. This procedure is used only when the intended residues are not present in the force field. In our case, this approach was taken to derive parameters for: s-2,3-diamino propanoic acid (Dap), 4-(aminomethyl)benzoic acid (Amb) and for a lysine residue able to form peptide bonds with both amino groups.
4	Build a molecular mechanics topology for each peptide dendrimer, i.e., a set of parameters defining all the atoms, bonds, angles, charges and potential places for interactions, for each and every atom in the dendrimer. In other words, the set of all the bonded and nonbonded parameters characterizing a given molecule.
5	Create a unit-cell around the dendrimers; impose periodic boundary conditions; chose a model describing the solvent molecules (waters) and add the solvent to the unit-cells.  **Legend:** Stereo Image of the C1 dendrimer surrounded by 9442 water molecules. The simulation box has the shape of a rhombic dodecahedron. The water molecules are colored blue and represented as lines while the dendrimer is represented with "sticks" colored by atom - H(gray), O(red), N(blue), C(green), S(yellow).
Simulation Stage: Energy Minimization	
6	Perform an initial energy minimization consisting of 12 000 steps using the steepest descent algorithm with all the dendrimer heavy atoms positionally restrained.
7	Perform 15 000 steps of a second energy minimization also with the steepest descend algorithm, but without any position restraints.
Simulation Stage: Initiation	
8	The initiation procedure starts with a 20 ps MM/MD simulation with all dendrimer heavy atoms restrained, at a temperature equal to 298.15 K and assigning initial random velocities to the atoms.
9	In a second step, another MM/MD simulation runs for 30 ps restraining only the alpha carbon atoms (T = 298.15 K).

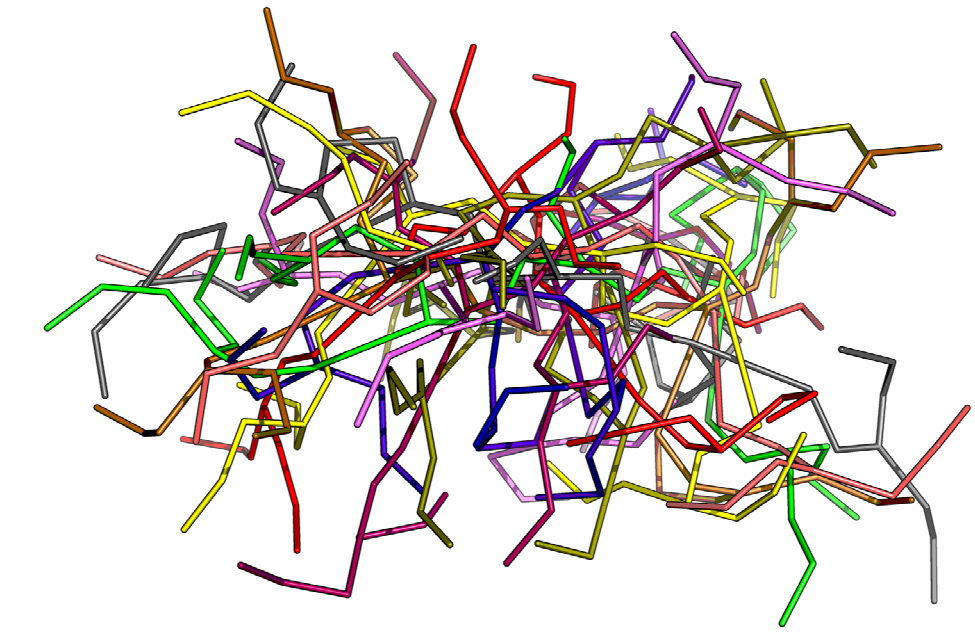
Continues on next page

10	Perform a 50 ps simulation without position restraints and while heating the system up to 400 K.
11	Change the Coulombic charges of each dendrimer atom from their reference value to a positive one (+0.1 *e*), thus promoting repulsion among atoms. Continue heating through a simulation during 100 ps and up to 500 K.  **Legend:** 3D Structure of the B1 "hedgehog" conformation. The coordinates where recovered 110 ps after the initiation stage begun. The picture on the left presents all the atoms except the aromatic and aliphatic hydrogens. The atoms are colored by element - H(gray), O(red), N(blue), C(green), S(yellow). The picture on the right represent a "skeleton-structure" of the peptide dendrimers, i.e., a line is drawn between each consecutive alpha carbon atoms. Also present is the line connecting the nitrogen atom of the amino cap and the subsequent alpha carbon. From both representations, it is perceivable that by changing the atomic charges in the course of the simulation we impose a repulsion effect on the atoms and the dendrimers "stretch".
12	Run a simulation during 50 ps decreasing the systems temperature to 400 K while keeping the positive charges on the atoms.
13	Maintain all the atoms positively charged throughout a 50 ps simulation while cooling the system down to its reference temperature (298.15 K)
14	Set the charges of all the dendrimer atoms to their reference values and complete the initiation protocol with a 20 ps MM/MD run at 298.15 K with all dendritic heavy atoms positionally restrained. At the end of this step we have the conformation from which to start the Equilibration/Production simulations.

Continues on next page

Simulation Stage: Equilibration and Production
15

Perform a MM/MD simulation for 100 ns at a temperature of 298.15 K, a pressure of 1 atm, using a time step of 2 fs and saving the atomic coordinates at every 10 ps. The distinction among equilibration and production phases is done on a later stage based on the analysis of statistically relevant properties.



Legend: Ten overlapping 3D conformations of the B1HHH dendrimer. The atomic coordinates were obtained from the trajectory of the first replicate by collecting the data at 10 ns intervals. This picture intends to represent some of the dynamic fluctuations occurring during the simulation time. Each of the ten conformations is represented using a different color and through the use of a line drawn between each consecutive alpha carbon atom. Also present is the line connecting the nitrogen atom of the amino cap and the subsequent alpha carbon.

Figure 2.4 – Step-wise description of the work performed

Steps (1) to (7) were performed only once for each of the five dendrimers in study (see Table 2.1) while steps (8) to (15) where repeated ten times (for each dendrimer) using different initial velocity values. Accordingly, we have performed a total of fifty standard MM/MD simulations, each 100 ns long.

2.2.11. Generation of MD input files and parameters

Before a simulation can be performed it is essential to select an initial configuration for the system. As aforementioned for inhomogeneous systems comprising a solute molecule immersed in a solvent, the starting conformation may be obtained from an experimental technique such as X-ray crystallography or NMR, or may instead be generated by theoretical modeling [Leach 2001].

Although much effort has been put into it, up until now no experimental 3D structure of a peptide dendrimer has been obtained. Therefore, we had to build the desired dendrimer through computational modeling approaches. The parameters we have considered for the topology of our peptide dendrimers are already contained in the GROMOS96 53A6 set (see below).

Designing a three-dimensional atomic structure of a molecule is a task that has been greatly simplified in the past years due to enhancements in the available software. In our case we have used the currently available free version of PyMOL [DeLano 2002] (version 0.99rc6) to obtain an initial set of 3D coordinates for each peptide dendrimer depicted in Table 2.1.

We have developed a systematic method to generate dendritic topologies by using and manipulating PyMOL libraries with some in-house scripts.

PyMOL contains by default an atomic 3D description of some proteinogenic amino acids in its fragment libraries. However, some of the residues presented in our structures had to be created, manipulated and added to PyMOL's libraries. This is the case of the acetyl caps, the amino cap, the 4-(aminomethyl)benzoic acid (Amb) residues, the s-2,3-diaminopropanoic acid (Dap) branching residues and of a lysine residue specially altered to form peptidic bonds through both the amino groups therefore serving as a branching unit (from this point forward, we will refer to this "special" lysine as *Lyr*, meaning *branching lysine*).

The fragments/residues we have created are presented next in Figure 2.5 and the names assign to each of their atoms are highlighted.

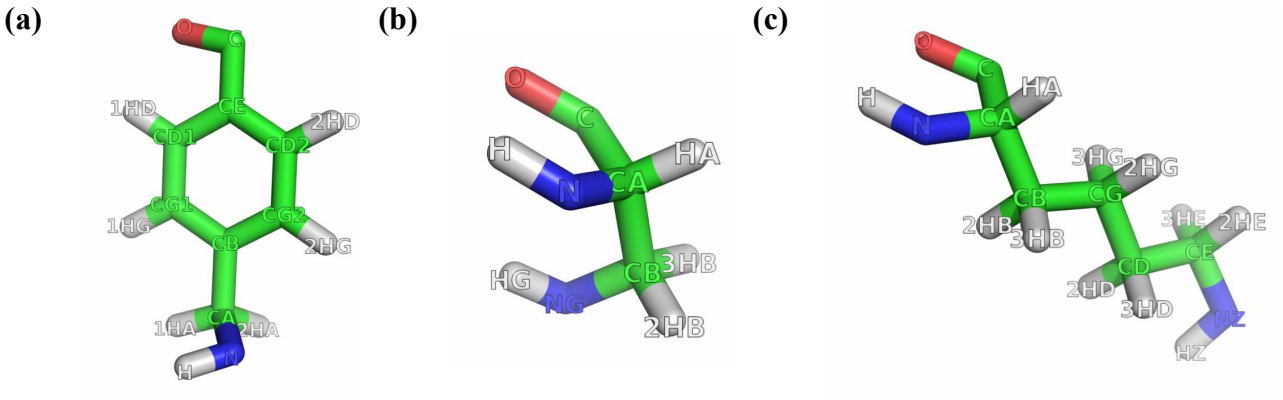


Figure 2.5 – Atomic composition of the residues designed and parameterized

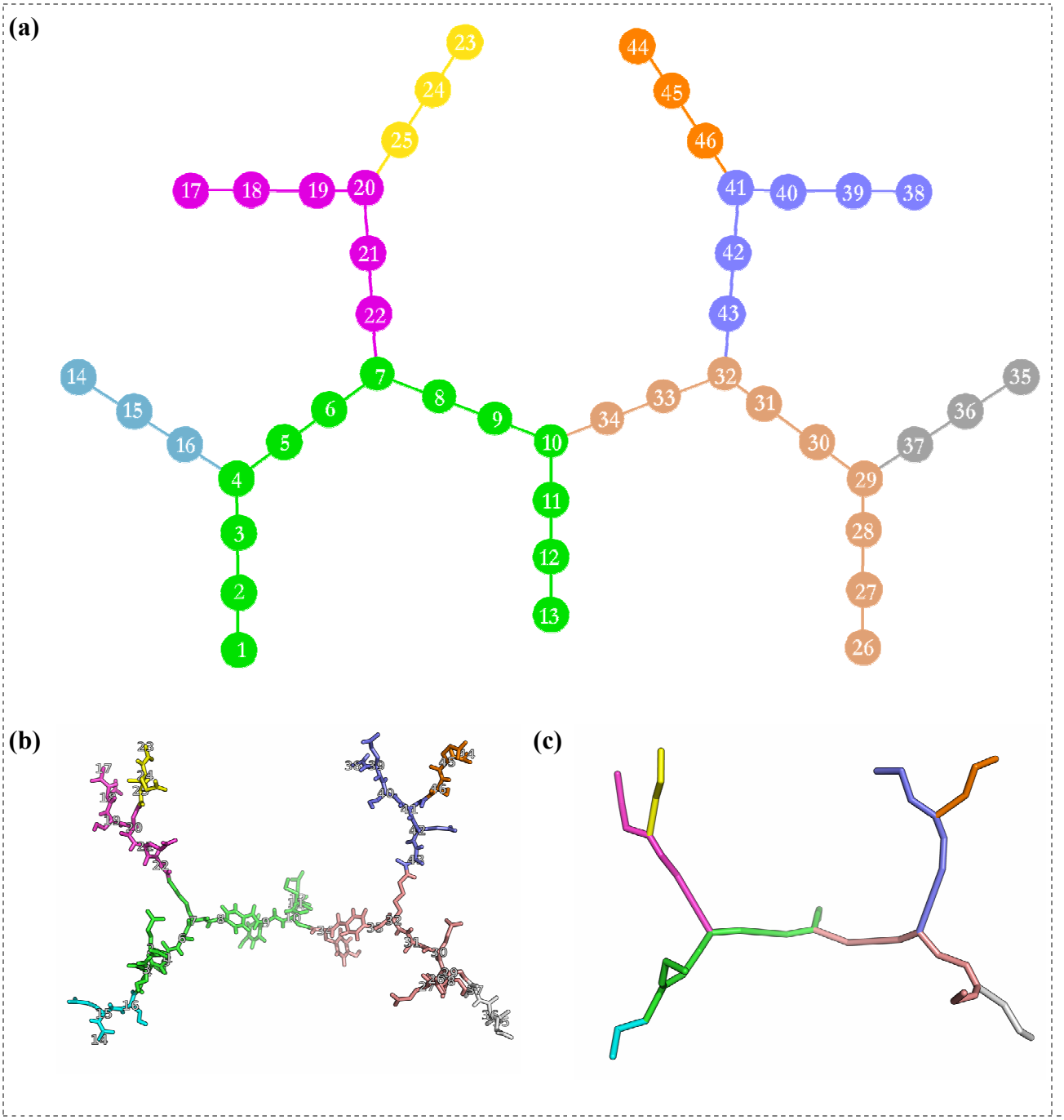
Images representing the atomic structure of the fragments designed using PyMOL and parameterized for the GROMOS96 53A6 force field. **(a)**4-aminomethyl(benzoic) acid, Amb; **(b)**s-2,3-diaminopropanoic acid, Dap; **(c)**Branching lysine, Lyr.

The atoms are colored by element - H(gray), O(red), N(blue), C(green).

The reader should pay particular attention to the nitrogen atoms, specially NZ(Lyr) and NG(Dap) because these are the atoms what will allow for the occurrence of bifurcations and are the debouch points of the nascent chains.

Contrary to proteins and other linear polymers, the dendritic topology does not present an obvious and unambiguous starting or ending point for numbering the different residues. Therefore, we have defined a convention for the numbering of peptide dendrimers. The idea underlying our choice for the numbering of residues is to maximize the number of "true" peptidic chains[11]. This convention is depicted in Figure 2.6.

[11] By "true" peptidic chains we mean chains presenting a main chain with the usual configuration found in peptides and proteins.



Continues on next page

(d)

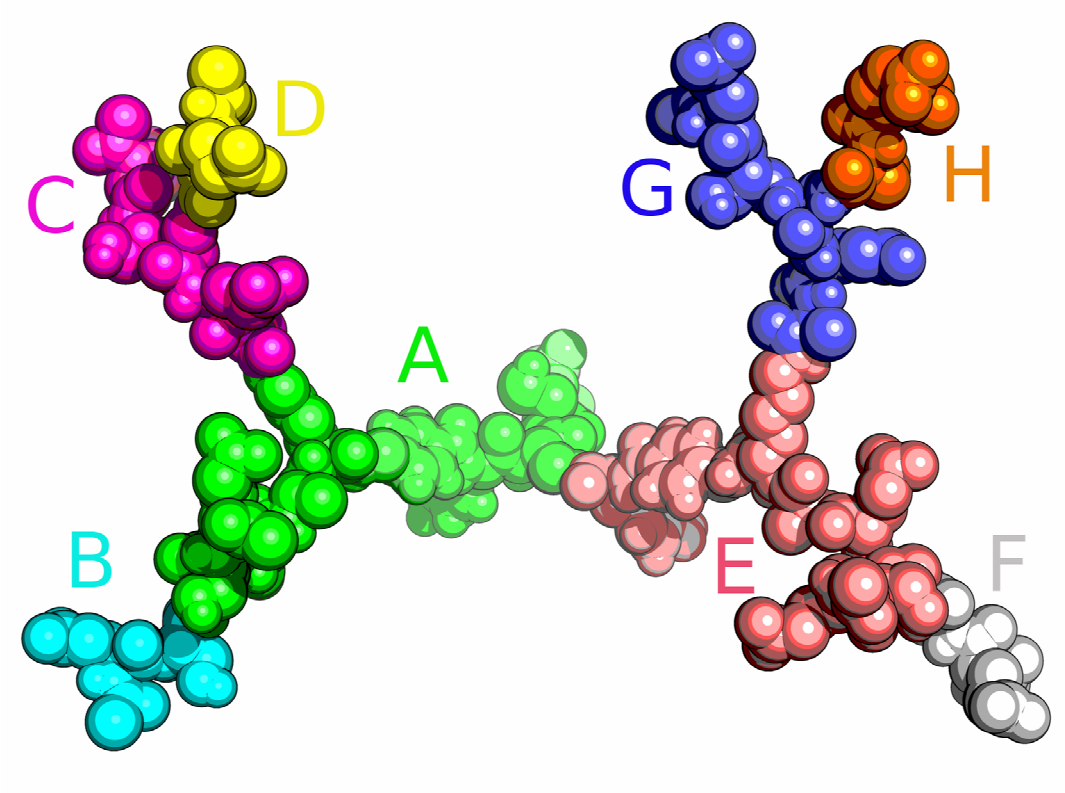


Figure 2.6 – Numbering and labeling of peptide dendrimer residues and chains

Images illustrating the convention defined for numbering of the different residues and the name, start-point and end-point of each chain in the dendrimer.

(a) Schematic representation of all the chain and the order with which the residues are numbered. The numbering proceeds always from the periphery to the core. Each branching, spacer and cap unit is represented by a circle.

(b) Three-dimensional image of the B1 dendrimer in "hedgehog" depicting all atoms other than the aliphatic and aromatic hydrogens. The atoms are colored according to their chains and the residue number is presented superimposed on each of the alpha carbon atoms. The enumeration of residues is exemplified by the B1 dendrimer but holds its validity for any of the dendrimers in study.

(c) Three-dimensional image of the B1 dendrimer in "hedgehog". Each chain is represented using a different color and through the use of lines drawn between each consecutive alpha carbon atom. Also present is the line connecting the nitrogen atom of the amino cap and the subsequent alpha carbon.

(d) Three-dimensional image of the B1 dendrimer in "hedgehog" using spheres to represent each atom. The dendritic chains colored according to the names we have assign them.

As can be ascertained from the Figure 2.6 the first element to be numbered belongs to the outermost residues (in this case an acetyl fragment) and the corresponding chain (chain A) proceeds until the core, ending in the core capping group, specifically the amino group. The second chain (chain B) starts at the outermost residue (also an acetyl group) adjacent to the first one and ends in the common branching unit. This numbering logic is repeated until all the residues are accounted for, and in the end we will have as many chains as peripheral end residues. Thus, every new chain will end with a bond to a NZ or NG atom of the previous chain (see Figure 2.6(b) and Figure 2.7). The branches (chains) were constructed using an antiparallel β-sheet orientation [Cantor 1980; Leach 2001].

The convention applied for the numbering of atoms and residues is important because it will allow us to define three-dimensionally the different configurations adopted by dendrimers along the simulations.

The dendrimer architecture described is schematically illustrated in Figure 1.4 (see section 1.9.). The structure emerges from a bifunctional core and emanates radially outward with bifunctional branching at every two spacer residues.

In Figure 2.7 we emphasize that all chains, other than the first, have a main chain composed by, at least, one “false” peptide bond. This means that the carbon atom involved in the peptide bond is connected to either a NG (in case the branching unit is Lyr) or a NG (in case the branching unit is Dap) atom instead of the usual N atom.

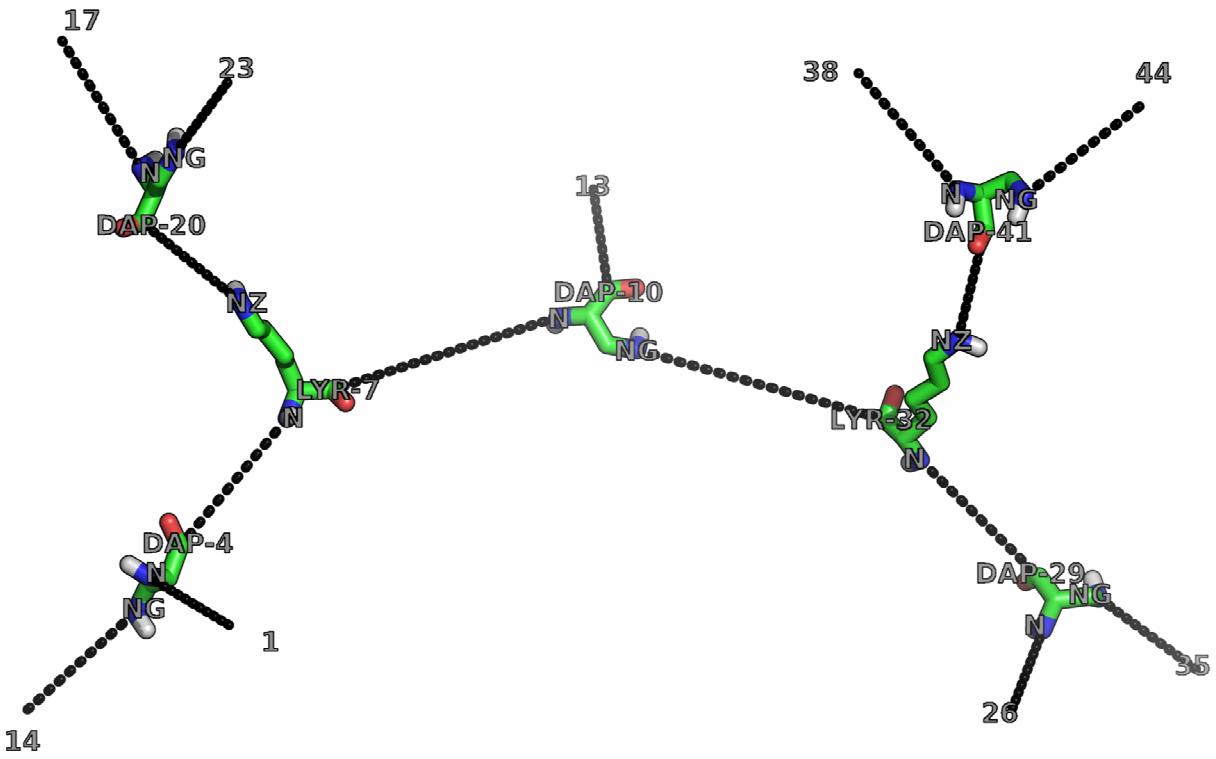


Figure 2.7 – Schematic representation of the branching positions within a dendrimer

This scheme illustrates the B1 dendrimer topology. The branching units are labeled with the residue name and number. Also indicated are the names of the atoms responsible for the establishment of peptide bonds and the number of the peripheral and core caps. We emphasize that all chains, other than the first, have a main chain composed by at least one "false" peptide bond.

Detailed descriptions of the branching units are presented in Figure 2.5 From that figure we can conclude that Dap is a much more symmetric branching unit than Lyr. This is important and will be extensively discussed in the next chapter.

The starting structures obtained with PyMOL showed configurations resembling a random coil (see step 1 in Figure 2.4). As such, previous to the Production stage of the simulations, (step 15 in Figure 2.4) all structures were subjected to thorough minimization and initiation procedures, thus ensuring that no conformational state is privileged as a consequence of the starting 3D conformation.

Once the model structures are selected, the topology information (e.g., connectivity, atomic types, partial charges, etc.) must be input into GROMACS and the necessary parameters supplied to perform the initial calculations.

All the information included in the description of the force field is static; it is never modified during the MM/MD runs.

This is a crucial step, initiated by identifying molecules already present in the force field that closely mimic the residues that compose our dendrimer and that are not accounted for in the default force field set up.

Most of the compounds that compose our dendrimers are already parameterized, the exceptions being: Dap, Amb, Ace, Nh2 and Lyr.

The GROMOS96 53A6 force field parameters for these residues were constructed manually by adapting some of the existing residues in the parameter files. These items topologies are described with detail in Appendix B. The charges assigned to each tritable residue during the MM/MD simulations match the ones typically present at pH 7.

The number of water molecules used in each peptide dendrimer simulation is given in Table 2.2. Other relevant informations are also provided.

Table 2.2 – Relevant informatiom regarding the simulations performed

Somes details on the conditions used to simulate each peptide dendrimer. The number of water molecules present in each system is directly related to the size of the simulation box. [a]Number of residues that constitute each peptide dendrimer. Althought the caps are ubiquitous and their definition as "true" residues is doubtful, they were included in this count; [b]Number of atoms that constitute the peptide dendrimer; [c]Total amount of water molecules present in each of the simulated systems; [d]Initial distance between the dendrimer atom closest to the end of the box and the end of the box; [e]Branching residues topology.

Dendrimer	Residue number[a]	Atom number[b]	Water molecules[c]	Box distance[d]	Branching[e]
B1	46	380	13619	2.500	5 dap, 2 lyr
B1H		386	13608	2.415	
B1HH		378	9446	1.752	
B1HHH		384	9441	1.878	
C1		368	9442	1.879	7 dap

As can be ascertained from Table 2.2 the systems initial conditions are quite identical. All dendrimers present thirty seven residues and a general architecture were each two spacer residues are intercalated with a branching one (see Table 2.1).

From the previous table, one can also conclude that the total atomic size of the systems under study varies between 30 and 40 thousand atoms. All the systems are mostly composed by water, having about 10 thousand water molecules and only a dendrimer one.

This huge amount of water is intentional and, although it will considerably slow down the computational time, it is essential to ensure meaningful results. The fact is that previous to the analysis of some of our simulation results we knew nothing about this peptide dendrimers conformational flexibility and compactness. As a matter of fact, given the state of the art articles on dendrimers, all sort of behaviors could be expected (see Chapter 1). Consequently we choose a unit-cell with dimensions (and associated number of water molecules) assuring that each dendrimer representation does not interacts with its periodic counterparts, i.e., periodic images are always separated by a distance (and a number of waters) greater than the cut-off for electrostatic interactions, 1.4 nm.

2.2.12. Equilibration and Production

The MM/MD familiarized reader has certainly noticed that in the sketch of the simulations protocol presented (see Figure 2.4) there is not a clear distinction between the Equilibration and Production phases. As a matter of fact only through a first analysis of the results we were able to decide at which time each simulation was equilibrated (see the next chapter).

This approach was adopted because we are dealing with systems to which the time scales of the processes leading to significant fluctuations are unknown and therefore we have to choose and test different measures before we can select one that fully describes/monitors the system's fluctuations. Accordingly, the different systems and replicas showed to be equilibrated at different time lengths and only the equilibrated trajectories were used.

2.3. Article-like Outline

The GROMACS package [Hess 2008; Lindahl 2001; Berendsen 1995], version 4.0.2, and the GROMOS96 53A6 force field [Oostenbrink 2004] were employed to perform MD simulations on five different peptide dendrimers. The starting structures were obtained by theoretical modeling using PyMOL [DeLano 2002], version 0.99rc6, and some in-house scripts. All peptide dendrimers were designed and simulated with an amine cap on the C-terminus and acetyl caps at the N-termini. All residues were charged in the MD simulations which can be regarded as representing pH 7. Ten molecular dynamics simulations were performed for each of the peptide dendrimers in study, for a total of 50 simulations (5 dendrimers, 10 replicates each), each 100 ns long.

The replicates of each peptide dendrimer started from the same optimized system but with different sets of random velocities. All simulations were done with explicit solvent solvating each peptide dendrimer with an adequate number of single point charge (SPC) water molecules [Berendsen 1981] in rhombic dodecahedral unit-cells, applying periodic boundary conditions. Suitable minimum distances between the peptide dendrimers and the end of the unit-cells were used. The final systems contained about 40 to 30 thousand atoms.

The equations of motion were integrated by the Verlet Leap-Frog algorithm using a time step of 2 fs, and the system's coordinates (snapshots) were save every 10 ps for further analysis.

The LINCS algorithm [Miyamoto 1992] was employed to keep all bonds at their equilibrium values and the SETTLE algorithm was used for keeping water molecules rigid.

The non-bonded interactions were treated with a twin-range cutoff of 8/14 Å and updated every 10 fs. The reaction-field method [Robertson 2008; Tironi 1995], with a

relative dielectric constant of 54.0 [Smith 1994], was used for the long-range electrostatic interactions. Neighbor lists were updated every 10 fs.

Solvent and solute were separately coupled to temperature baths at 298.15 K, with Berendsen coupling [Berendsen 1984] and relaxation time of 0.1 ps. A Berendsen isotropic pressure coupling was used at 1 atm, with relaxation time of 0.5 ps and isothermal compressibility of $4.5 \times 10^{-5} bar^{-1}$.

All systems were energy minimized to remove excessive strain. An initial energy minimization was done consisting of 12 000 steps using the steepest descent algorithm (all dendrimer heavy atoms were position restrained) followed by another 15 000 steps of minimization using the same algorithm but without position restraints.

The initiation procedure consisted of seven different simulations. First a 20 ps simulation was performed with all dendrimer atoms restrained at a temperature equal to 298.15 K and assigning initial random velocities to the structure, thus generating the different replicates, followed by another simulation for 30 ps restraining only the alpha carbon atoms. A third simulation is performed for 50 ps without position restraints and heating the system up to 400 K. After that the Coulombic charges of each atom were changed from their reference values to a value of +0.1 *e* and heating continued during 100 ps up to 500 K.

Next a simulation was run during 50 ps decreasing the systems temperature to 400 K while keeping the positive charges on the atoms. All the atoms were kept positively charged throughout a 50 ps simulation while cooling the system down to its reference temperature (298.15 K). Finally, we set all the charges to their original values and completed the initiation protocol with a 20 ps MD run at 298.15 K with all dendrimer heavy atoms positionally restrained. The force constant used for position restraints during the minimization and initiation steps was 1000 $kJ.mol^{-1}.nm^{-1}$.

Starting with these fully extended configurations, each replicate (10 per dendrimer) was then simulated without position restraints for 100 ns. The systems showed to be equilibrated at different time lengths and only the equilibrated trajectories were used.

Chapter 3

Results and Discussion

In the former chapters we have described in detail the biological-physical problem we are interested in, and the best way to study it. Still, the work itself is worth (in a very pragmatic view) by the results it provides and by the scientific-based conclusions it generates. Quoting Becker and Watanabe: "Clearly, the value of any computer simulation lies in the quality of the information extracted from it" [Becker 2001].

The analysis of molecular trajectories resulting from MD simulations will attest the quality of the assumptions and models used. Furthermore, it is this analysis of trajectories that will allow us to know and understand the characteristic processes of the system.

Over the next topics we will discuss a wide range of properties of the simulated systems in the attempt to find the ones that typify them. As it has been explained in the first chapter, in this initial phase of peptide dendrimer MD studies, one is more interested in the "what" than in the "why".[12] With this work we hope to shed the foundations for future investigations.

---

[12] One might consider this as an exploratory work, in the search for the "right angle" to approach peptide dendrimers.

When required we will introduce theoretical concepts fundamental to the analysis approach used.

3.1. Validation and Equilibration

As in any study, a very important issue to deal with is validation. Before we proceed to any other type of analysis this issue must be addressed to ensure the results obtained are relevant, and to assess whether the simulation methods and force fields employed are accurate enough to model peptide dendrimers.

The overall quality of a simulated system and corresponding properties will depend on different aspects [van Gusteren 1998]: first, the quality of the model used; second, the ability of the potential energy-function functional form and parameters to accurately reproduce the dynamic behavior of the system; third, the size of the systems used, is it sufficiently large to avoid distortion of the dynamics by the unphysical spatial boundary conditions?; and fourth, is the necessary degree of sampling, statistics, and convergence reached?

In this context, a useful criterion to validate our model is to determine some quantities that can be compared to experimental values. Given the lack of structural and dynamic information on these particular peptide dendrimers, this poses as a challenge.

In the article published by Reymond and coworkers [Sommer 2008] on B12-binding dendrimers, they presented the values for the hydrodynamic radius (R_H) and diffusion coefficient (D) of the B1G2 dendrimer (i.e., the B1 dendrimer but only up to the second generation – $(AceGluAla)_4(LysAmbTyr)_2DapCysAspNH_2$).

Luckily for us in their unpublished data Reymond and coworkers performed these same measurements for the B1 dendrimer as a whole, and they were kind enough to share the results with us. The values are presented in Table 3.1. These values were

obtained by diffusion NMR spectroscopy for both the free and the aquocobalamin complex forms. To the best of our knowledge, these are the only dynamic quantities so far experimentally measured measured that can also be calculated by averaging over an ensemble of configurations of our MD sets.

Table 3.1 – Experimental diffusion coefficients and hydrodynamic radii for B1 free and complexed forms in solution [Sommer 2008]. The values where obtained by Reymond and coworkers through diffusion NMR spectroscopy.

	D (m^2 s^{-1})	R_H (nm)
$B1_{free\ form}$	1.0833×10^{-10}	1.840
$B1_{aquocobalamin\ complex}$	1.0845×10^{-10}	1.838

Herein, we will focus our efforts on the diffusion coefficient since its determination does not depend on assumptions concerning the molecule shape[13]. Furthermore, the diffusion coefficients are sensitive to structural properties of the observed molecular species such as weight, size, and shape, as well as binding phenomena, aggregation, and molecular interactions [Cohen 2005].

For each replicate of each peptide dendrimer the total number of hydrogen bonds (#H-Bonds), the solvent accessible surface area (SASA), the radius of gyration (R_g), the sum of the distance among the branching units alpha carbons (Distance α-C), the minimum distance with the periodic image (Minimum Distance), and the distance between the two farthest atoms (Maximum Distance), in the dendrimer were calculated. This way we expect to identify the property, from the previous set, that more accurately captures the "nature" and variability of the systems.

[13] Contrary to the diffusion coefficient, the hydrodynamic radius, or Stokes-Einstein radius, can not be easily calculated without assuming that a molecule presents a well defined geometrical shape, normally a sphere. Furthermore, the hydrodynamic radius can not be calculated without the diffusion coefficient value.

Each of the aforementioned properties is obviously interesting and provides valuable results by themselves, and they will be thoroughly discussed further in this thesis. Nevertheless, we want to identify the one that, by comparison to all the others, is more adequate to perform an initial description of the systems.

For an entire MD simulation, as well as for separate blocks from the simulation, examining the convergence of a property, A, can be most rigorously carried out by calculating the time series[14] of that property and determining its autocorrelation function and accompanying relaxation time [Becker 2001].

From the analysis of the mentioned time series the initial portion of the simulations during which relaxation of A occurs can be identified and discarded (equilibration phase). Additional analysis and convergence verifications are performed only on the remainder *production simulation* [Becker 2001].

The property, A, we are "searching for" is, therefore, the one whose variation occurs in a time scale significantly smaller than the length of our simulations. Furthermore, the simulation time should be much longer than the *relaxation time*[15] (t_{relax}) of the property considered.

In Figure 3.1 we present scatterplots [Everitt 2001] combining different pairs of the previously mentioned properties, for all 100 ns long replicates of the B1 dendrimer. These plots will allow the visualization of the "dominant" property.

[14] Calculating any dynamic variable, $A(t)$, along the trajectory obtained in our MD simulations results in a "time series". In the present context time series means a sequence of data points correlated and spaced at an uniform time interval.

[15] The relaxation time corresponds to the time a properties takes to demonstrate an equilibrium behavior. The relaxation time of a certain property will depend on the type of property, the thermodynamic state and the type of molecule [van Gusteren 1998].

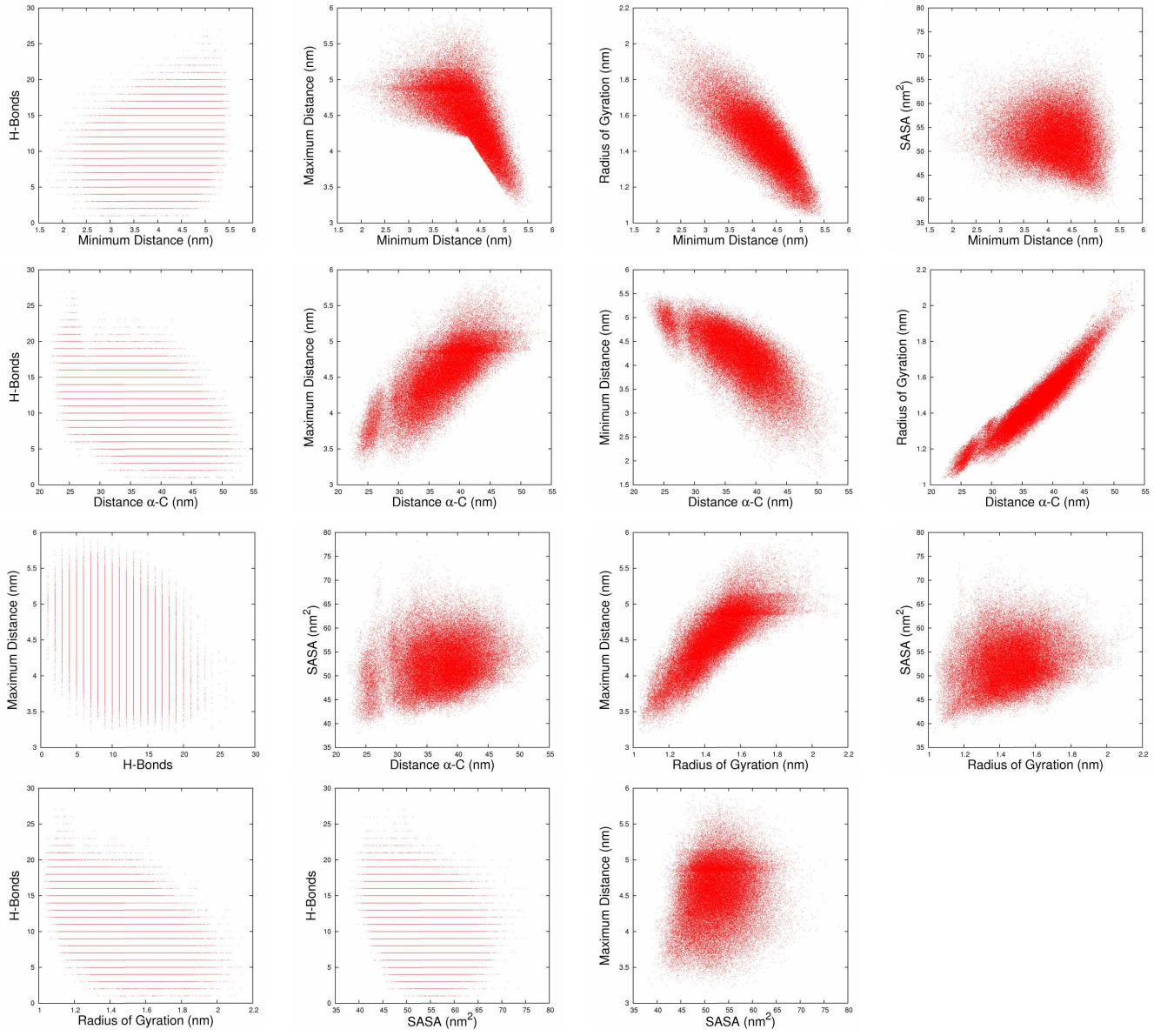


Figure 3.1 – Scatterplots combining multiple properties calculated for B1.

Scatterplots displaying the relation between pairs of properties computed for the B1 peptide dendrimer. The values were determined using all replicates and the entire 100 ns trajectories. The properties considered are: the total number of hydrogen bonds (#H-Bonds), the solvent accessible surface area (SASA), the radius of gyration (R_g), the sum of the distance among the branching units alpha carbons (Distance α-C), the minimum distance with the periodic image (Minimum Distance), and the distance between the two farthest atoms (Maximum Distance).

The corresponding plots for the other dendrimers under study can be found in Appendix C.

From these plots, we note the predominance of the radius of gyration, mainly because it conditions the dispersion of points when plotted against all the other properties.

Moreover, R_g is a property that presents small variations (see Figure 3.2). Therefore, the equilibration of the systems was monitored through the radius of gyration, R_g. A brief description of the radius of gyration concept and calculation is thereby mandatory.

The *radius of gyration* can be considered as a rough measure for the compactness of a structure [van der Spoel 2005(b); Leach 2001; Allen 1989]. It is a commonly used measure to define branched structures [Lee 2006; Lee 2002; Karatasos 2001; Han 2005] and can be conceived as a quantitative estimate of a dendrimer size.

For each simulation the R_g was calculated according to:

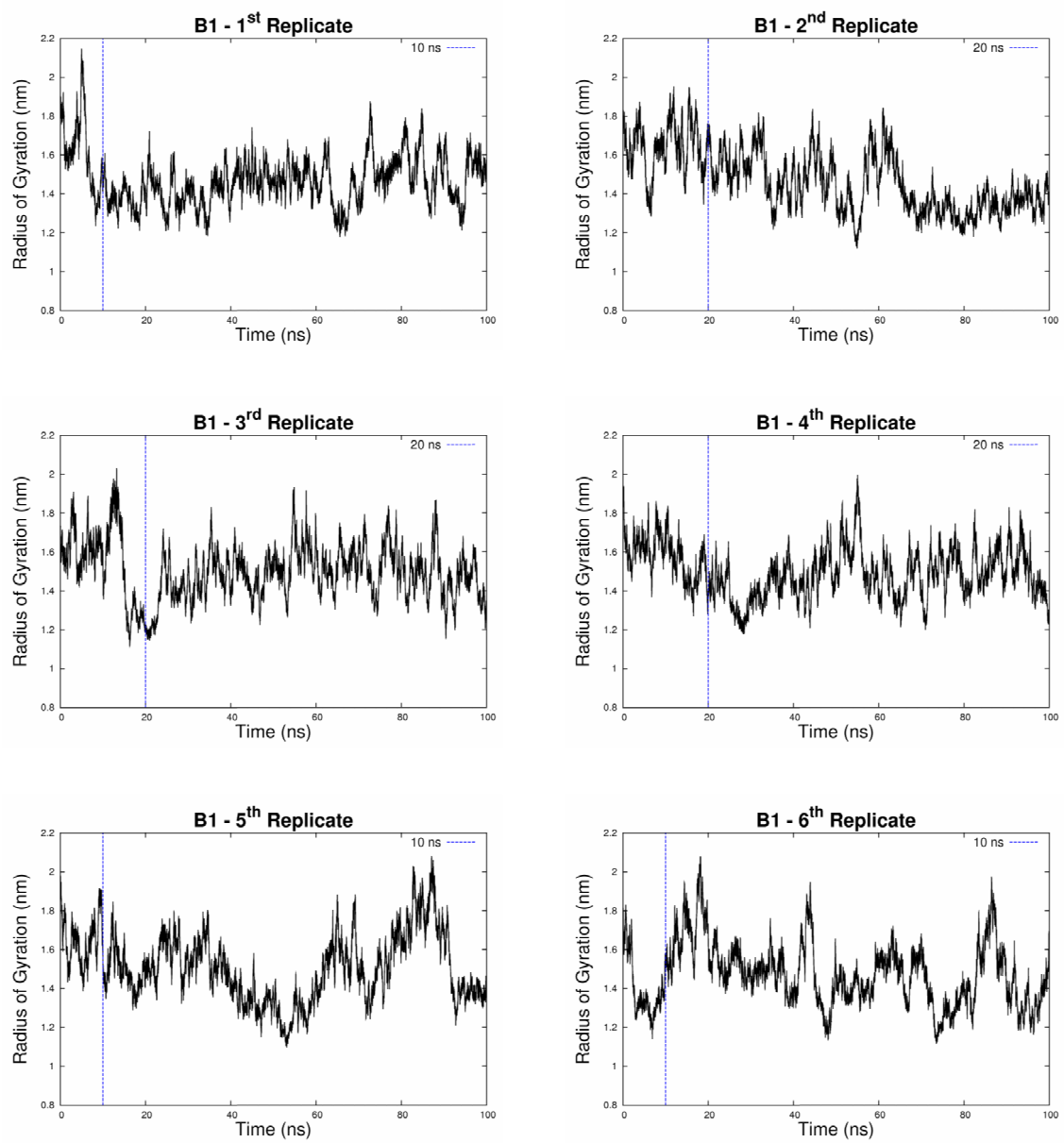
$$R_g = \left(\frac{\sum_i \|\mathbf{r}_i\|^2 m_i}{\sum_i m_i} \right)^{\frac{1}{2}}, \qquad \text{Equation 3.1.}$$

where m_i is the mass of atom i and $\mathbf{r}_i$ the position of atom i with respect to the center of mass of the molecule.

The relaxation and dynamics of the different properties occurring in a MD simulation can be analyzed in different ways, but they all must combine favorably to yield reliable results.

When starting a simulation from a non-equilibrium initial state the rate of relaxation towards equilibrium for different properties will give an indication of their relaxation time. In addition, the fact that a certain property is in fact at equilibrium can be confirmed/monitored by the time series of the property, its average and its autocorrelation function [van Gusteren 1998; van Gusteren 1995].

The most straightforward analytical procedure to define the equilibration interval for each of the simulations is to plot the R_g time series. Such plots are presented in Figure 3.2 for B1 (the plots for the other dendrimers can be found in Appendix C) giving a quick and easy overview of the simulation; these plots are especially useful for picking up trends or sudden transitions from one state (or configuration) to another.



Continues on next page

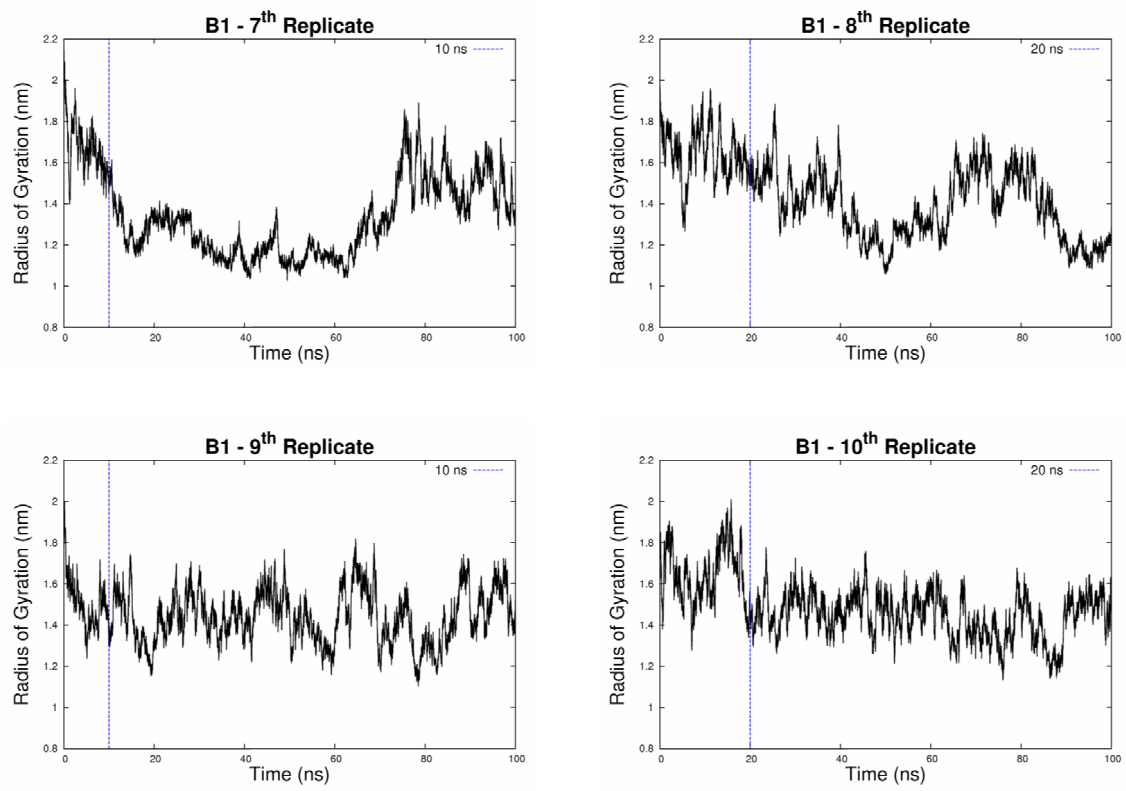


Figure 3.2 – Radius of gyration time series for B1.

Values of the radius of gyration plots as a function of time for all ten replicates. Each plot corresponds to a different replicate. The 100 ns trajectories are depicted using points collected every 10 picoseconds. A vertical line is drawn where each replicate is consider to be equilibrated.

One may be concerned with the length of our simulations because the timescale for equilibration of these sorts of systems is not known. By studying the trends exhibit by the radius of gyration curves along time for each replicate we can decide the duration of the equilibration period for each one of them. As each replicate departed from a different state[16] they will achieve equilibrium at different times. Furthermore, the average value for the equilibrium radius of gyration of each replicate can give us an idea of the "level of sampling" accomplished with our simulation. In other words, if

[16] A different initial velocity, see the previous chapter.

simulations of the same system departing from different initial conditions converge to similar average conformational spaces then our simulation time is enough to allow an efficient search of the conformational space.

Table 3.2 presents the equilibrium times for all the replicates of all the simulated systems and also the average value for the R_g of each replicate considering only the equilibrate values, i.e., a particular time interval at the beginning of the 100 ns long simulation.

Table 3.2 – Equilibration times for each dendrimer's replicates (time removed from 100 ns trajectories); and average radius of gyration for each equilibrated replicate.

Replicate	B1		B1H		B1HH	
	t_{Equil}[a] (ns)	<R_g>[b] (nm)	t_{Equil}[a] (ns)	<R_g>[b] (nm)	t_{Equil}[a] (ns)	<R_g>[b] (nm)
1	10	1.4556	5	1.4589	10	1.4540
2	20	1.4296	5	1.4498	10	1.4244
3	20	1.4867	10	1.4502	10	1.5227
4	20	1.4774	10	1.3341	15	1.4277
5	10	1.4968	15	1.4761	10	1.3533
6	10	1.4777	5	1.4651	5	1.3057
7	10	1.3062	10	1.4887	5	1.4254
8	20	1.3810	10	1.3246	10	1.4565
9	10	1.4391	15	1.4743	10	1.4829
10	20	1.4421	10	1.5292	10	1.4696

Replicate	B1HHH		C1	
	t_{Equil}[a] (ns)	<R_g>[b] (nm)	t_{Equil}[a] (ns)	<R_g>[b] (nm)
1	10	1.4165	15	1.0536
2	10	1.4701	10	1.1544
3	10	1.4477	20	1.0641
4	10	1.4623	20	1.0816
5	20	1.3036	20	1.0225
6	10	1.5207	10	1.1190
7	10	1.4516	30	1.0389
8	10	1.3585	20	1.0710
9	15	1.3504	10	0.9949
10	15	1.3762	10	0.9931

[a] Amount of time after the beginning of the 100 ns long simulation that the system takes until achieve equilibrium.

[b] Average radius of gyration value of the equilibrated trajectory.

Thus, by concatenating the trajectories of the different replicates we obtain the total amount of production simulation in equilibrium conditions for each system.

Table 3.3 – Total length of the concatenated production simulation for each dendrimer.

Dendrimer	Total Equilibrated Simulation Length (ns)
B1	850
B1H	905
B1HH	905
B1HHH	880
C1	835

In MD simulations and given the statistical nature of the method itself, one can never be absolutely sure that convergence has been achieved. Still, efforts can be made to ensure that the properties one intends to study, such as R_g, have reached satisfactory levels of convergence. From the previous table we can deduce that all simulations "effectively converged", meaning that after a certain amount of time the R_g values for different simulations (replicates) converge, and as such, the equilibration times chosen are adequate for the present conditions and the present property.

The minimum time considered for equilibration was 5 ns and the maximum was 30 ns (from runs with total lengths of 100 ns). By analyzing these values we can conclude that the variations described by the radius of gyration are effectively small and that the structural transitions occurring in peptide dendrimers take a great amount of time when compared with other molecular systems such as peptides or proteins.

The decay time of the autocorrelation function will give us one idea of t_{relax} for R_g, thus allowing us to conclude the maximum time span for the structural shifts associated with that specific property.

The measure of the relaxation of the peptide dendrimer molecules is determined by their *time autocorrelation functions*, *C(t)*, for the property that "best" describes the system, in this case R_g, whose $C_{Rg}(t)$ is given by:

$$C_{R_g}(t) = \frac{\langle R_g(t) R_g(0) \rangle - \langle R_g \rangle^2}{\langle R_g{}^2 \rangle - \langle R_g \rangle^2} .$$ Equation 3.2.

The details on the equation components and correlation functions in general can be found in Appendix D.

This function measures the correlation of property R_g to itself at two different times, separated by a time interval (Δt), averaged over the entire trajectory.

In principle, the average over all the replicates in the simulation should yield a better representation of the equilibrium properties for each system. As such we have calculated the correlation function for each of the equilibrated replicates, but also the average time autocorrelation function. The autocorrelation functions for the individual replicates are presented in Appendix E.

The replicate-averaged autocorrelation functions are represented in Figure 3.3. The determination of the average autocorrelation function for the radius of gyration is not an obvious mathematical process, and therefore, the equation used for the calculation of the average time autocorrelation function is presented in Appendix D.

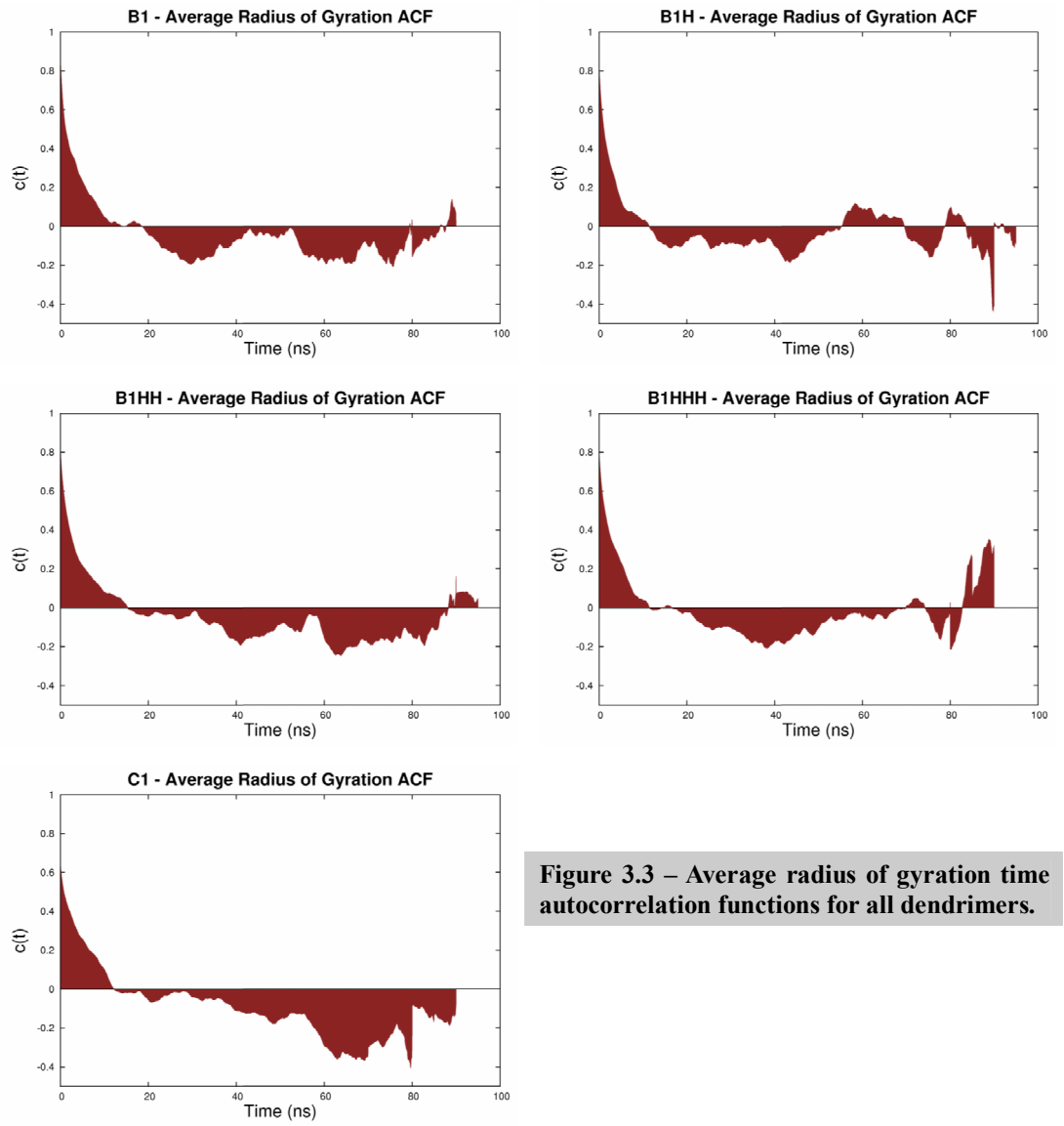


Figure 3.3 – Average radius of gyration time autocorrelation functions for all dendrimers.

The physical behavior of peptide dendrimers will be reflected in the statistical properties of the autocorrelation functions previously presented. The autocorrelation function of R_g for a succession of values implies that the system as a finite "memory".

This means that the system starts with a particular R_g,at an initial time, and after a certain amount of time, the system will lose memory of its initial state. That time interval is named *correlation time* (t_{corr}).[17]

[17] Due to the fact that if two conformations in a MD trajectory are comprised in that time interval they can be considered as time correlated.

The radius of gyration correlation times for the peptide dendrimers under study range from 10 ns (B1H) to 20 ns (B1), being here defined by the time interval that the autocorrelation function curve takes to cross the x-axis (the first root value or "zero" of the function).

A correlation time of approximately 15 ns for these systems suggests that equilibration is accomplished within the simulated time scales.

According to the plots presented in Figure 3.2 and Figure 3.3 the relaxation and equilibration times are on average five times less than the total time of each simulation. This suggests that the total simulation time is long enough to sample the peptide dendrimers conformation space and to compute averages over static and dynamic properties. Furthermore, when we concatenate all the equilibrated trajectories, or average a property over all replicates, we are considering a trajectory that is approximately 850 ns long (see Table 3.3).

Therefore, our simulation times seem to be enough to properly sample the conformational space of equilibrated dendrimers.

Given the fact that the radius of gyration is a property with a slow time span (approximately 15 ns) for structural variations in this systems, and that it also captures the intrinsic variability of these systems better than any of the other properties studied, we can expect all the other properties (solvent accessible surface area, number of hydrogen bonds, among others) to be equilibrated after the interval considered for R_g.

An overview of relaxation behavior, of sampling and convergence used to decide the validation and equilibration of MD simulations are presented in detail elsewhere [van Gusteren 1998; van Gusteren 1995].

In addition to what was previously mentioned on the equilibration of dendrimers, it is important to emphasize a practical aspect concerning the simulations performed; the minimum distance between the periodic images. Therefore, we must confirm that the size of the system is sufficient to avoid interactions among dendrimers placed on different "simulation boxes".

In Appendix F are shown plots representing the time series of the minimum distance between periodic images for all dendrimers.

From those plots it is evident that in the course of the simulations the distances among periodic images are always greater than the cutoff for the electrostatic interactions (1.4 Å). Therefore, the periodic images do not "see" each other. Moreover, one can also observe from the minimum distance plots that the dendrimers start in an extended conformation and gradually relaxe to a more equilibrated and less extended one.

We can now return to the discussion of diffusion coefficients and pursue the validation of our models by comparison of calculated and experimentally observed properties.

As previously mentioned, the parameterization of a new set of molecules demands its validation against experimental data, in our case the diffusion coefficient, *D*.

Computational approaches for the calculation of transport coefficients (such as *D*) from MD simulations include: numerical evaluation of equilibrium correlation function; observing Einstein relations (Equation 3.3.); or conducting non-equilibrium MD simulations [Berendsen 2007; Allen 1989].

Next, we present the equation for calculating the diffusion coefficient from equilibrated MD simulations using the Einstein relation.

The diffusion coefficient [Berendsen 2007; Allen 1989] of a single-molecule is given by[18]:

$$D = \frac{1}{3}\int_0^\infty \mathrm{d}t \left\langle \mathbf{v}_i(t) \cdot \mathbf{v}_i(0) \right\rangle, \qquad \text{Equation 3.3.}$$

where $\mathbf{v}_i(t)$ is the center-of-mass velocity of the molecule.

The results obtained will be nearly identical to averaging over all atoms [van der Spoel 2005(b)] composing the molecule.

[18] Considering all three dimensions of the Cartesian space.

The correspondent Einstein relation, valid for times much longer than the correlation time of the velocity, is:

$$6Dt = \left\langle \left\| \mathbf{r}_i(t) - \mathbf{r}_i(0) \right\|^2 \right\rangle, \qquad \text{Equation 3.4.}$$

where $\mathbf{r}_i$(t) is the molecule position. We note that, in practice, an important aspect of determining D using Equation 3.4., is that all calculations must be performed for a single molecule, this is, it is important not to switch attention from one periodic image to another.

Diffusion coefficients for B1 in water were computed from the mean-square displacement (MSD) using the Einstein relation [Allen 1989]. For this, we determine the MSD for each replicate and did a linear fit to the linear part of an average-replicate MSD curve.(see Figure 3.4).

To be on the "safe side", we have considered a margin of 20 ns for equilibration of all replicates, thus disregarding the values presented on Table 3.2 and placing all replicates in equal terms of number of points used to calculate the mean-square displacement (MSD) curve. This assumption also helped greatly in the determination of the average MSD over all replicates.

We were sure to ignore the first 20 ns because it is the time that the slower replicate of B1 took equilibrate according to the observation of R_g (see Table 3.2). Thus, we chose a time where all replicates seemed equilibrated.

As a consequence of the previous considerations, all final trajectories considered were 80 ns long.

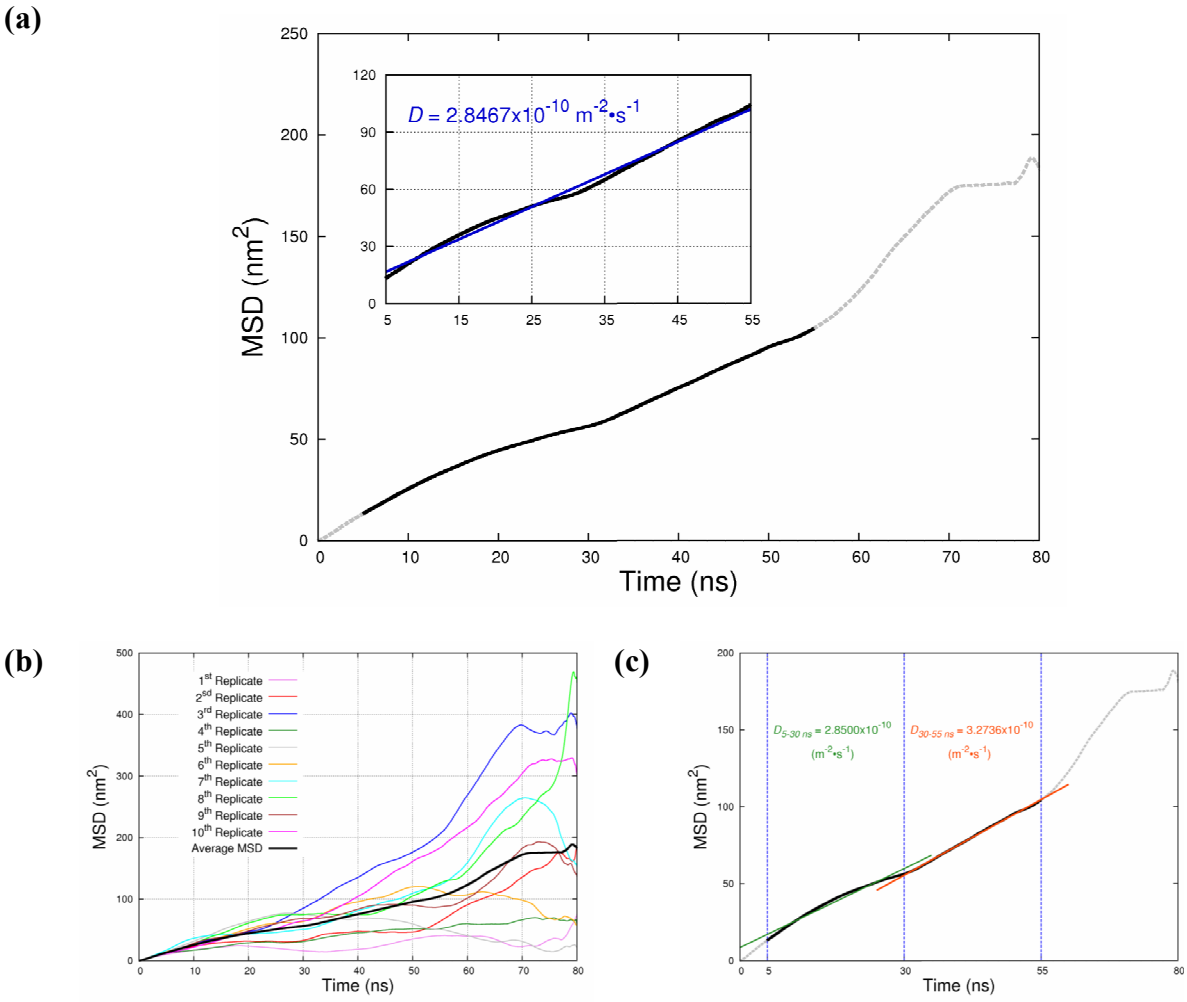


Figure 3.4 – Determination of the diffusion coefficient for B1.

(a) Curve drawn by averaging over all the MSD values of the different replicates; the dashed part of the curve represents the time values that were neglected in the fit. The fit over the considered interval are present in the adjacent box, along with the diffusion coefficient value obtained. **(b)** The MSD evolution for each of the replicates in the time interval considered (from 20 to 100 ns, in a total of 80 ns for each replicate); the average MSD over all replicates is evidenced as the black curve. **(c)** The intervals and linear fits considered for the determination of the diffusion coefficient associated error. The error is calculated as the difference of the diffusion coefficients obtained from fits over the two halves of the fit interval (see the text for details).

The choice of the interval over which one fits a straight line to the MSD curve (see Equation 3.4.) determines the value of D.

Hence, as the initial part of the MSD represents ballistic motion[19] it is preferable to exclude some of the initial points from the calculation of the least squares fit whose slope yields the diffusion coefficient [Chitra 1997]. Also, the sampling is poor at larger times, and it is advisable to discard these points as well [Chitra 1997].

Accordingly, we have decided to use only the points in the range between 5 ns and 55 ns. A least squares straight line was obtained by fitting to the points of the MSD curve over this range. Figure 3.4(a) depicts the curve drawn by *MSD* as a function of time and the fitted line.

The Einstein's relation yields a diffusion coefficient of 2.8467×10^{-10} m^2 s^{-1}. The estimate of *D* obtained is higher then the value obtained experimentally (listed in Table 3.1). We consider this to be the best estimate of *D* due to the judicious choice of the range over which the straight line has been fitted and the averaging over all ten replicates each contributing with a sampling equivalent to a 80 ns long simulation.

As we intend to compare the diffusion coefficient obtained by MD with its experimental homologous it is of considerable importance to know the relative accuracy of the value we have determined. Therefore, we determined the diffusion coefficient associated error. A significant drawback of most simulation approaches concerns the difficulty in obtaining error estimates associated with some calculated properties (for example, specific heat, isothermal compressibility, among others). This happens because these quantities depend on fluctuations in related quantities. The diffusion coefficient, in particular, is even more difficult to evaluate accurately because it deals with time-dependent as well as fluctuation-dependent quantities [Chitra 1997].

Methods to estimate the errors involved in the results obtained from simulation have been discussed in the literature [Allen 1989; Becker 2001]. In the present work, we used a method described in [Micaelo 2006]. The error is calculated using the difference between the diffusion coefficients obtained from fits over the two halves of the fit interval of the diffusion coefficient. We have considered two separated ranges: the first from 5 ns to 30 ns, and the last from 30 ns to 55 ns.

[19] Ballistic motion refers to a “free” path motion characterized by many degrees of freedom.

We fitted two separated straight lines over the points of the MSD in these ranges and obtained values of $D_{5\text{-}30ns} = 2.8500\text{x}10^{-10}$ m^2 s^{-1} and $D_{30\text{-}55ns} = 3.2736\text{x}10^{-10}$ m^2 s^{-1}, respectively (see Figure 3.4(c)). The error value is $\pm 4.2690\text{x}10^{-11}$ m^2 s^{-1}.

As a concluding remark, we must empathize that the simulation results are qualitatively correct, with both the modeled, (2.8467 ± 0.4269) $\text{x}10^{-10}$ m^2 s^{-1}, and experimental ($1.0833\text{x}10^{-10}$ m^2 s^{-1}) values in the same magnitude, although the theoretical diffusion coefficient is overestimated, probably due to the lack of sampling for such a dynamic and fluctuating property.

When comparing the computational and experimental values available for the diffusion coefficient some considerations are important. We are in fact comparing a value determined for a single molecule that freely diffuses in an homogeneous solvent (water), without further contact with other molecules, with a value obtained through an experimental technique[20] where the observed values result from an ensemble of configurations that may interact with each other, thereby imposing several restrictions to each others' diffusion. The problems affecting the values obtained for *D* using MD are out of the scope of the present thesis, but a detailed discussion can be found in an article by Chitra *et al* [Chitra 1997].

In conclusion, judging from the comparison with the experimental data on diffusion coefficient for these molecules, the simulation adequately reflects the experimental system. Furthermore, an agreement between simulation and experiment exists, and properties obtained by further analysis will hold validity as predictive properties. Thus, the set of simulations implemented can be regarded as a stringent test to the transferability aptness of the GROMOS96 53A6 force field do deal with peptide dendrimers, corroborating the former validation [Javor 2009].

[20] In this particular case Reymond and coworkers used Diffusion NMR Spectroscopy to obtain their results [Cohen 2005].

3.2. Conformational Analysis

3.2.1. Introduction

This point in the text is probably a nice place to summarize what as been done so far. Thus far we have recognized that the folding of a dendrimer in water cannot be directly followed across time with a single MD simulation (or other of the available simulations methods). As such we have sampled the conformational states of each dendrimer using MD, and employed the ergodic hypothesis to consider a number of separated conformations as an ensemble that reflects (we hope) the multiple folding substates available to a dendrimer.

We now must analyze the ensemble, whether by characterizing its individual properties, or by calculating its free energy as a function of a number of folding-progress coordinates.

Under the definition of folding previously presented[21], it becomes of crucial importance to understand the relation between coalescence phenomena, compactness, and structure acquisition as folding proceeds (*folding dynamics*).

The folding of macromolecules is better understood (and explained) in the context of free energy surfaces [Guo 1997; Udgaonkar 2008; Frauenfelder 1998]. According to this theory, folding can better be described by the descent of the folding molecule down a multidimensional folding funnel[22], with local roughness reflecting the potential for transient trapping minima, and the overall slope of the funnel representing the drive to the *native state* [Guo 1997].

---

[21] Folding does not refers to a progressive pathway of unique single conformations, but rather to interconvertions in a back-and-forth progression from the unfolded to folded state [Maisuradze 2009].

[22] The funnel perspective of folding implies that the process proceeds through a continuum of intermediates where there is no single sequential route [Udgaonkar 2008; Frauenfelder 1998].

3. Results and Discussion

The perspective of folding as the climbdown of an energy landscape is a very integrative view of the problem, because it characterizes the different possible conformations, but also determines the dynamics of the system [Frauenfelder 1998; Becker 1997(a); Maisuradze 2009; Brooks 2002].

An important question that is crucial to answer before we go further in the text is: in the context of peptide dendrimers what are effectively their folded conformations?

At this point of the text this question might seem strange, but the reader must consider the following reasonable doubts: is there a state we can call the *native state*? How do we define it? How do we define the less "ordered" structures? As *molten globules*? As *random coils*? Most probably, the common elements of protein secondary structure (e.g., β or α patterns [Cantor 1980]) do not exist, so, how do we define the collective arrangement of the different branches? These are all important questions we must answer before continuing. The definitions we will employ here are based on the descriptions used for proteins, thus, some are generally applicable. However, other definitions were particularly thought to fit specific aspects of dendrimer folding (as the *bowl-like* term).

We start by discussing a concept with which the reader is probably familiarized, the *native structure* (or *state*). The common denomination of native structure (that is tightly associated with folding) refers to the well defined, functional, three-dimensional structure of a molecule (mainly proteins). However, this definition is unsatisfactory because it does not recognize the existence of molecules that are known to present their highest functional rates at high disorders conformational states. Transposing these ideas to peptide dendrimers, we verify the impossibility of defining their *native structure* based on functional aspects. Thus, in here we will employ a broader definition of the *native structure(s)* encompassing the global minimum on the free energy surface for the molecule.

So, to understand and characterize the folding of peptide dendrimers we must, in practice, identify the accessible energy minima on the potential-energy surface, thus finding not only the *native conformation* (global minimum), but also the other *conformational substates* (relative minima).

Despite the computational power made available to scientists nowadays, scrutinizing the conformational variability of complex biological systems through

visual inspection has become an infeasible task with the currently accessible sampling techniques. Accordingly, strategies allowing for high throughput data analysis have gained relevance as tools for the selection and description of the most representative conformations.

Conformational analysis[23] is the study of the conformations[24] of a molecule and their influence on its properties [Leach 2001], being a major computational tool used to probe the structures and flexibility of molecules. This approach is employed to investigate, not only the most frequent and stable structures (if they exist), but also the overall flexibility and energetic dependence of molecules [Becker 1997(b); Leach 2001]. The imperative drive of conformational analysis is something some authors call the *conformational search* [Leach 2001], the objective of which is to identify a smaller subset of characteristic low energy conformations.

This approach works under the assumption that the *preferred conformation* are associated with energy minima in the correspondent energy surface. As such, the understanding of dendrimer folding requires locating the three-dimensional structures they can adopt at the minima points of their energy landscapes.

In the physical world, conformations are gathered based on kinetic transitions occurring within the corresponding *energy basins*[25]. An ideal conformational analysis procedure will detect conformations belonging to the same energy basin and distinguish between conformations that are part of different energy basins

---

[23] The assessment of the relative energies (or thermodynamic stabilities), reactivities, and physical properties of alternative conformations of a molecular entity, usually by the application of qualitative or semi-quantitative rules or by semi-empirical calculations. IUPAC Definition [Moss 1996].

[24] The spatial arrangement of the atoms affording distinction between stereoisomers which can be interconverted by rotations about formally single bonds. Some authorities extend the term to include inversion at trigonal pyramidal centers and other polytopal rearrangements. IUPAC definition [Moss 1996].

[25] "A *basin* corresponds to the set of conformations whose steepest descent path along the energy hypersurface leads to a particular minimum, establishing a one-to-one relation between basin and minima; in short, all conformations that 'fall' to the same minimum belong to the same basin" [Campos 2009; Becker 1997(a)]. Or in other words, energy basins are regions of structures connected by low energy barriers.

(characterize by high energy transitions/barriers) [Becker 1997(a); Becker 1997(b); Leach 2001].

Some studies demonstrate that reaction coordinates are correlated with the overall features of the folding process [Shea 2001]. Hence a detailed conformational analysis must start by establishing the most suitable reaction coordinates. Therefore, the results obtained will always be a function of one or more of these reaction coordinates. A number of reaction coordinates may be used to describe the folding of a macromolecules. In the so called *protein science* one normally uses the number of native contacts, the root-mean-square-deviation or the radius of gyration as a reaction coordinate [Guo 1997; Campos 2009].

The main criterion supporting the choice of a reaction coordinate is that it unambiguously distinguishes between the folded and the unfolded/partially-folded conformations [Shea 2001; Campos 2009].

To extract the conformational properties of the peptide dendrimes that are being studied, the conformational ensembles that were sampled must be analyzed. This analysis will focus on global properties, attempting to characterize features such as overall flexibility or to identify common trends in the conformation set.

It should be stressed that the different conformational analysis tools used can be applied to any collection of molecular conformations. Nevertheless, the study of dendrimers implies some extra tricks and "maneuvers", but these will be explicitly refereed in the appropriated topics.

A free energy landscape is defined as the "energy" of the system as a function of all its coordinates [Frauenfelder 1998]. For molecules in vacuum the "energy" is the potential energy, while for solvated molecules it is the solvent-average conditional free energy (or potential of mean force) [Campos 2009]. In molecular systems the nominal dimensionality is three times the number of atoms minus some degrees of freedom for rotation, translations and possible constraints [Hamprecht 2001]. Since it is neither possible (except for very small systems) nor desirable to represent the free energy surface as a function of all $3N$-6 coordinates of a molecule[26], we need to

[26] This corresponds to a ($3N$-6)-dimensional conformation space.

employ an approach that significantly reduces the system's degrees of freedom (and complexity), while satisfactorily representing its conformations.

Fortunately, the effective dimensionality of molecular systems is much lower than $3N$-6 dimensions, due to its physical and chemiscal constraints[27]. This legitimates that complex multidimensional systems can be described in low-dimensional spaces, and turns the study of their underlying folding processes from a herculean to "just" a hard task.

Based on the distribution of points (conformations) in these low-dimensional subspaces, a continuous estimate of the density of states can be obtained [Hamprecht 2001]. This implies an obvious relation between the sampling achieved throughout MM/MD simulations and the quality/detail accomplished by of the resulting free energy surfaces. A MM/MD trajectory can be viewed as a discretized path through a configuration space, with every configuration (or snapshot) corresponding to a single point in the path [Hamprecht 2001]. The different conformations thus obtained have the same chemical formula, but different structures and energies.

Different conformations sharing similar atomic properties are called *conformational substates*, and are usually separated by energy barriers.

It should be noted, however, that the term energy landscape is sometimes used for the multidimensional potential energy surface that underlies the molecule's conformation space, and sometimes for its free energy profile. In this thesis the focus is on the free energy profile and we will refer to it generally as an energy landscape.

[27] As an example of a dimensionality lowering factor consider the fact that two atoms involved in an unconstrained chemical bond will always be at a distance that corresponds roughly to their equilibrium bond length.

In equilibrated systems, the Gibbs free energy difference between two states (or conformations), is given by the natural logarithm of the ratio of their occupancies, K_{eq}:

$$\Delta G = -RT \ln K_{eq}, \qquad \text{Equation 3.5.}$$

with R the ideal gas constant and T the temperature of the system. The occupancies are, in turn, given by an integral over their densities in configuration space [Hamprecht 2001]. This means that from the density of states, the Gibbs free energy surface is known, which determines all thermodynamic observables. Using additional information more chemical properties of the system can be obtained. For example, the most favored conformations, which will correspond to the free energy minima [Becker 1997(a); Brooks 2002; Maisuradze 2009; Campos 2009].

The purpose of our study is to identify and characterize the most homogeneous conformational classes present in peptide dendrimers, using as few dimensions as possible to effectively capture the nature of the system hypersurface.

Our conformational analysis starts with a broad examination on the ability of certain conformational coordinates to preserve the structural similarity between conformations.

3.2.2. Representing peptide dendrimers

An important part in the analysis of molecular dynamics simulations is the visualization of structural two-dimensional examples depicting the features under study. An analogy with the visualization of proteins tells us that this is a hard task because macromolecules are complex structures, with multiple “components” that are sometimes packed closely and exhibiting different levels of structure.

Knowing this, and with the experience gained with 3D visualization of peptide dendrimers using appropriated software, we have selected a representation that makes the elucidation of the core and different branches position as intuitive as possible.

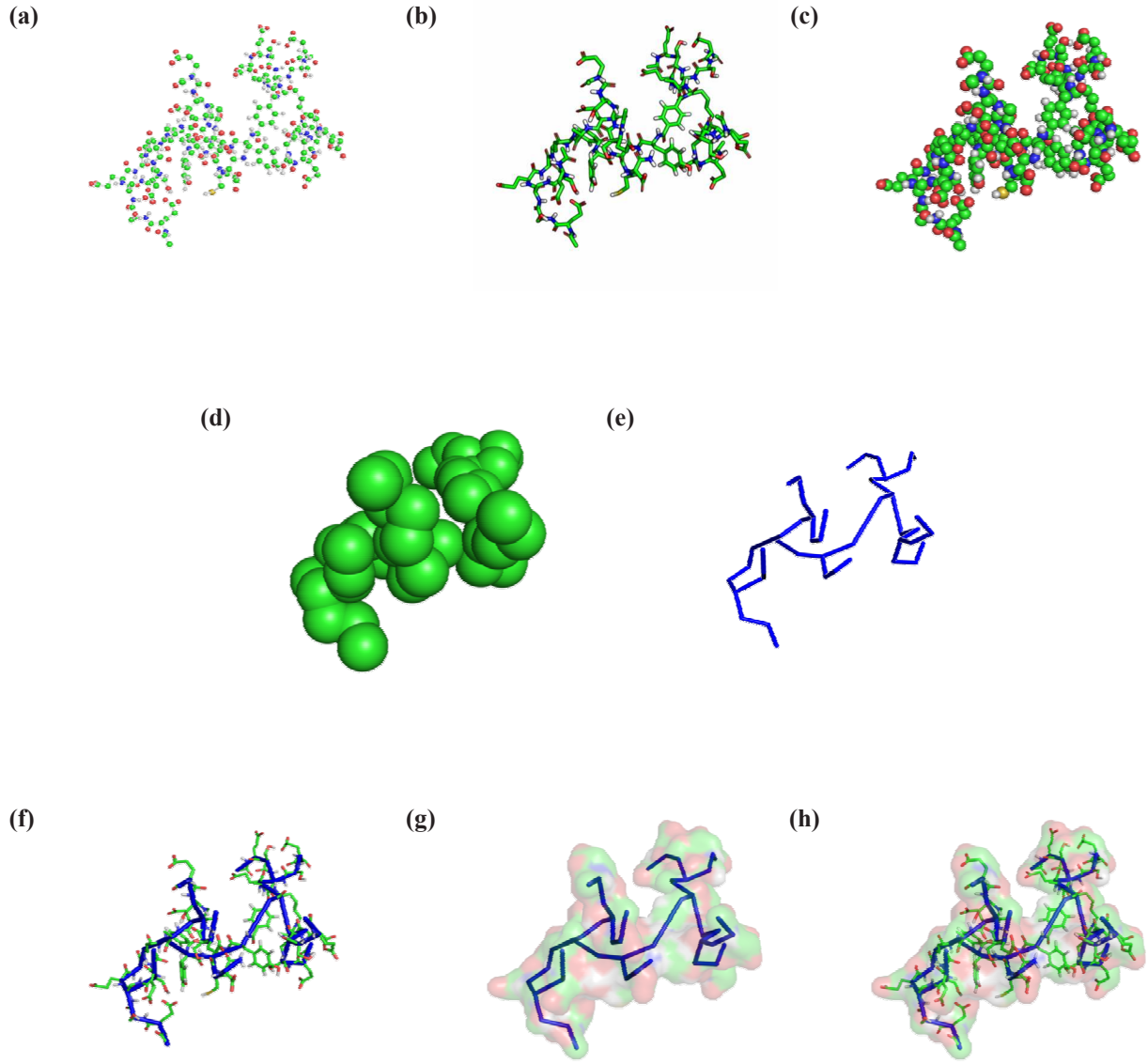


Figure 3.5 – Different representations of the same peptide dendrimer conformation.

All representation are oriented in the same view. The dendrimer conformation used belongs to B1HH; only the non-aliphatic hydrogens are represented. (a) Ball-and-Stick representation with all atoms colored by element. (b) All atoms and bonds represented as sticks and colored by element. (c) All atoms represented as balls and colored by element. (d) Only the alpha carbons are represented, but their are drawn as huge spheres, with the same size. (e) All residues represented considering only the alpha carbon atoms, and drawing lines between bonded residues (and between the nitrogen atom of the amino cap and the previous alpha carbon residue). (f) Representation of (e) in conjunction with (b). (g) Molecular surface representation combined with (e); surface is colored by element. (h) Representation of (g) in conjunction with (f).

As can be ascertained from the previous figure most representations of peptide dendrimers are “messy” and uninformative for the present purposes. In the present thesis we intend to investigate the conformational propensities of peptide dendrimers, as such, we need a representation that reflects the overall features of a conformation with minor “din”. Along the remainder of this thesis we will mostly represent them considering solely one atom from each of the constituting residues, namely the alpha carbon one (this is valid for all residues except the amino cap, were due to the lack of an alpha carbon atom we will consider the nitrogen one), and draw lines between atoms of residues that are in fact bonded (see Fig 3.5.(e)). This representation allows for immediate observation of the dendritic branch distribution and the core position. The main differences between miscellaneous conformations of the same structure can thus be easily and almost immediately envisioned.

Figure 3.5 also presents the schematic relation between our simplified description of the conformations, and the most descriptive ones. We consider that the simplified representation above described provides enough detail on the conformation of each peptide dendrimer without a major lost of detail (tacking into account the purposes and objectives of the present thesis).

Furthermore, from this point forward we will represent structures belonging to different dendrimers using a color code where: red is used for B1, green for B1H, blue for B1HH, yellow for B1HHH and gray for C1.

3.2.3. Histogram analysis

3.2.3.1. Radius of gyration

The analysis of MD trajectories can only be achieved through the study of the systems reaction coordinates[28]. As previously mentioned from a predefined set of properties we have selected the radius of gyration (R_g) as the geometric parameter that more suitably reflects the conformational variations occurring in the dendrimers along our simulations. The radius of gyration can be roughly considered as a measure for the compactness of molecular structures (see Equation 3.1.). The variations on the radius of gyration values are presented in Figure 3.2 and in Appendix C for the different dendrimers in study.

A careful discussion on this property seems wise as a first approach to investigate and identify the different conformations present and, if possible, group them.

From the plots in Figure 3.2 and Appendix C we can consider that the different replicates present values that are globally similar. This observation is important as an evidence of accurate sampling, supporting the sampling procedure employed.

[28] Reaction Coordinate: A geometric parameter that changes during the conversion of one (or more) reactant molecular entities into one (or more) product molecular entities and whose value can be taken for a measure of the progress of an elementary reaction (for example, a bond length or bond angle or a combination of bond lengths and/or bond angles; it is sometimes approximated by a non-geometric parameter, such as the bond order of some specified bond). In the formalism of "transition-state theory", the reaction coordinate is that coordinate in a set of curvilinear coordinates obtained from the conventional ones for the reactants which, for each reaction step, leads smoothly from the configuration of the reactants through that of the transition state to the configuration of the products. The reaction coordinate is typically chosen to follow the path along the gradient (path of shallowest ascent/deepest descent) of potential energy from reactants to products – IUPAC [Laidler 1996].

Another important consideration is that the dendrimers from the "B1-family"[29] (B1, B1H, B1HH and B1HHH dendrimers) present R_g values that vary roughly within a similar range of values. Still, there is a clear diversity in the R_g values of each replicate. Some trajectories (as the 9th replicate of B1, see Figure 3.2) show increases and decreases in R_g that vary significantly within a small interval; whereas other trajectories (as the 7th replicate of B1) exhibit little fluctuation between really distant values.

However, the C1 dendrimer, that presents the same global dendritic architecture but a very different residue composition, displays values of R_g that are inferior to the ones exhibit by the B1 family dendrimers. Moreover, the R_g values for C1 in equilibrium are more homogeneous among the different replicates than the ones from B1-family.

These intervals of values, both for C1 and the other dendrimers, implicate the existence of a more compact structure in C1 than in the other dendrimers. Exhibiting also a more stable behavior along the simulations (this conclusion can be ascertained merely form observing the plots of R_g as a function of time for each of the different replicates) and a smaller number of possible conformations visited along time (by comparison with the variations displayed in the same plots for the B1-family).

In the B1-family we observe that all dendrimers visit configurations with similar levels of compactness that seem to vary heterogeneously around a "central value".

Furthermore, and in contrasts with what happens in most C1 replicates, the B1-family dendrimers do not seem to adopt a conformation and remain stable, instead they seem to undergo a multiplicity of different R_g values in each different replicate.

An obvious implication from the plots in Figure 3.2 and Appendix C is that we are dealing with different kinds of behaviors, where some dendrimers can present more compact and "invariant" conformations than others.

[29] From this point forward we will consider a group of dendrimers – B1, B1H, B1HH and B1HHH – as a "family" in the sense that they share the same type of architecture, branching units and as mentioned in the Methods and Materials section a great number of identical amino acid residues placed in the same positions.

The calculation of the R_g values presented in the figures was obtained with GROMACS [Berendsen 1995] embedded tools (*g_gyrate* respectively) and some in-house scripts.

We collected a large amount of information regarding the R_g values adopted by different peptide dendrimers along a simulation. We can regard each of the R_g values obtained as a marked characteristic of a single geometric distribution of atoms at a precise time; this is the definition of a MD snapshot. Therefore, we will have as many snapshots as individual/specific conformations, and our goal is to identify and describe specific "classes" of conformations sharing a set of common overall features, i.e., conformations that present, when analyzing one particular (or a group of) reaction coordinate(s), a greater degree of similarity among them than with all other conformations. With the numerous conformations and associated R_g values we can start searching for preferences and trends.

We began by analyzing the distribution and frequency of the R_g values for all the different conformations. To perform this task we resort to histogram analysis, discussing not only the very informative shape of the histogram, but also depicting randomly picked snapshots of certain bins of the histogram to illustrate specific points.

For this analysis we have considered all coordinates of the systems obtained by concatenating the equilibrium trajectories of each replicate.

In a histogram the data is assign according to a segmentation of the total range of that data into equal sized non-overlapping intervals (bins or segments). We have chosen the size of the bins that more adequately reflects the aspects we intend to study.

In here we have found that the R_g values of the dendrimers in study always fluctuate between 0.8 and 2.2 nm, with variations accounted up to the third decimal place. As such we employed bins of 0.01 nm each[30].

The identification of conformational preferences and trends will be based mainly (at this stage) on the bin distribution. For instance, let us consider the hypothetical situation where there only exist closed (more compact) and open (more loose) conformations for each dendrimer, with a very small number of intermediate conformations; then one would expect that the histogram reflected this two phase behavior by drawing two separated peaks/lumps at different R_g values. The number of open or closed conformations would be expressed by the number of bins constituting each of the two peaks and their height. If this is our case, and we randomly select a set of representative conformations present in the same bin we can analyze whether the conformational variability within that bin reflects a trend, or if very different structures are found within the same R_g value, thus finding if this approach in general, and the R_g values in particular, are a satisfactory conformational descriptor/discriminator.

Over Figures 3.6, 3.7, 3.8, 3.9 and 3.10 we present the histograms obtained using the multiple R_g values, and we provide some images of the atomic geometries associated with some bins.

[30] The choice of a bin size does not obey strict criteria and we can not consider that if exists a "best" bin size or a "best" number of bin. Just an "optimum" been size that adequately reflects the trends "intrinsic" to the data that we want to study. Different bin sizes can in fact reveal different and miscellaneous features of the data, making the analysis of histogram results part subjective. In our work we have performed a trial-and-error iterative process to select the bin size that more properly expresses the conformational variability we expect to study with the histograms (data not shown).

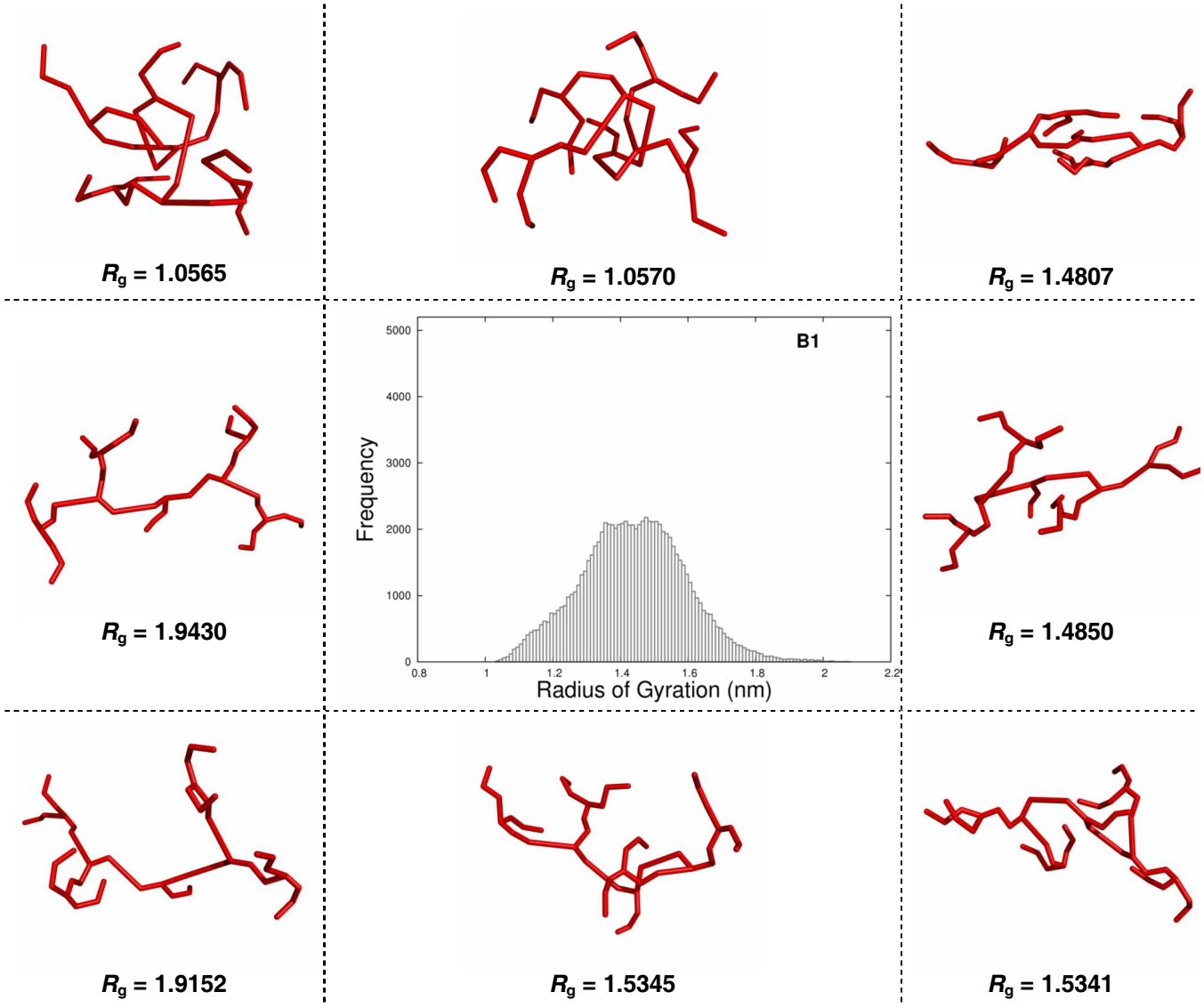


Figure 3.6 – Radius of gyration histogram for B1.

In the center of the figure we present an histogram constructed with all the R_g values collected from the equilibrated trajectories. Additionally we present atomic configurations chosen as examples of the populations within certain bins (see the text for further details).

The B1 dendrimer presents R_g values that vary between approximately 1 and 2 nm with most of the conformations presenting values between 1.30 and 1.55 nm, establishing a stable region in that range of values. The distribution of the data resembles a Gaussian curve but with a flatten peak.

From the shape drawn by the histogram with the adjacent pictures we verify that low R_g values (about 1.0 nm) correspond to very compact forms with some sort of compact nucleus formed by the core and some of the branches, while some of the other branches "float" freely. The number of branches floating and the number of branches that interact with the core vary (data not shown).

The most populated bins are characterized by a number of heterogeneous conformations with similar R_g values. The compactness of these structures is similar and looser then the one from structures with lower R_g values. The atomic geometries within the central bins are very different with no "global" organization pattern detectable.

Furthermore, within the most representative bins we can find structures with a more "ellipsoid" global shape (R_g = 1.4807) but also other with a more "bowl-like" shape (R_g = 1.5345).

Even for the structures displaying the highest R_g values we can not observe a marked trend, as demonstrated by the peripheral groups arrangement in the two examples shown. Still, we observe that structures with R_g values near 2 nm resemble the aforementioned "hedgehog" dendrimer conformation. Accordingly, we must consider that the conformational variability in B1 is huge, ranging from structures almost completely "open" (or stretched), displaying high R_g values, to more disordered states, with some compact areas and interacting branches (1.3 nm < R_g < 1.5 nm), and we can even find more compact states with multiple interacting branches.

Unfortunately this type of analysis does not allow us to identify different conformational trends for B1 (supposing they exist). Still, we must emphasize that these plots are too simple to account for most of the conformational space required to characterize peptide dendrimers. This problem was not surprising because peptide dendrimers are very flexible molecules and their trends are hardly described by a single reaction coordinate.

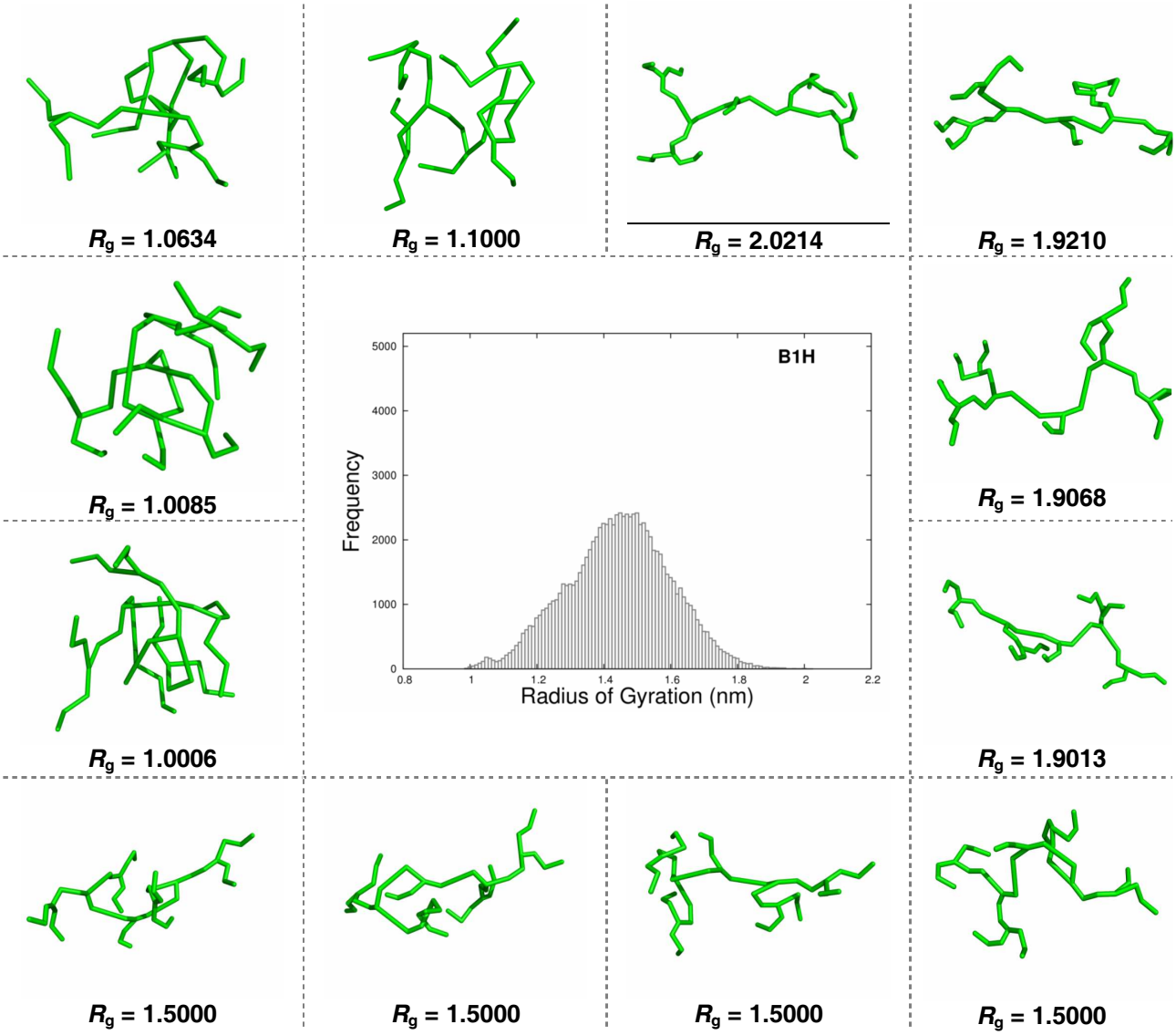


Figure 3.7 – Radius of gyration histogram for B1H.

In the center of the figure we present an histogram constructed with all the R_g values collected from the equilibrated trajectories. Additionally we present atomic configurations chosen as examples of the populations within certain bins (see the text for further details).

The B1H dendrimer presents values of R_g that also range between approximately 1 and 2 nm. Contrary to B1, the most representative R_g values are in the proximity of 1.5 nm and we can identify three important "lumps" in the data distribution. The first peak occurs between 1-1.1 nm, another between 1.3-1.4 nm and the "main one" near 1.5 nm. Conformations that have R_g values in the range of 1-1.1 nm are the most

compact ones and, as can be seem by the concomitant pictures, there exists a conformational heterogeneity within this peak, where some structures present three branches interacting near the core and one floating around it (R_g = 1.0634 nm), while others display all branches almost equally spaced in relation to the core and uniformly distributed (R_g = 1.0085 nm) in a very compact structure. As a consequence, we must conclude that similar R_g values can correspond to very different structures.

An interesting case illustrated in Figure 3.7 is the one where a randomly chosen value (1.5000) in the vicinity of the most frequent R_g value, numerous different conformations exist. These conformations are all "mildly open" with similar degrees of compactness, but the atomic dispositions vary.

For the conformations featuring the largest R_g values (1.9 nm < R_g < 2 nm) the situation is analogous to B1.

Next, we present Figures 3.8 and 3.9 with the histograms corresponding to B1HH and B1HHH respectively.

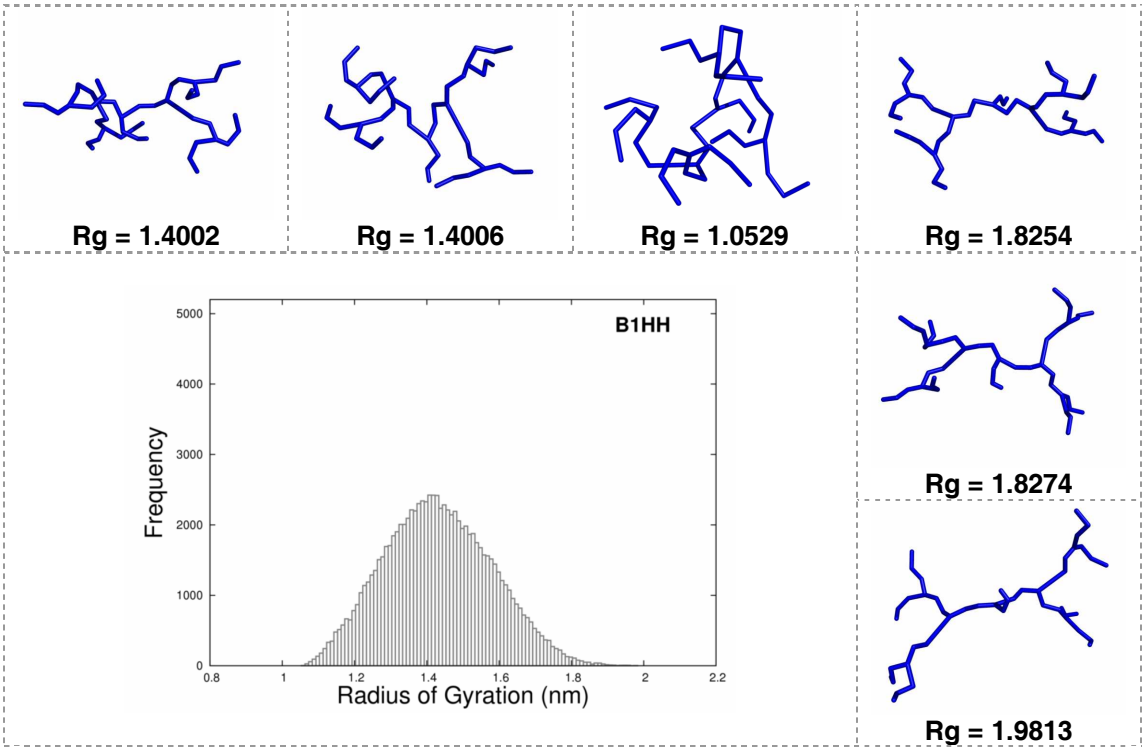


Figure 3.8 – Radius of gyration histogram for B1HH.

In the center of the figure we present an histogram constructed with all the R_g values collected from the equilibrated trajectories. Additionally we present atomic configurations chosen as examples of the populations within certain bins (see the text for further details).

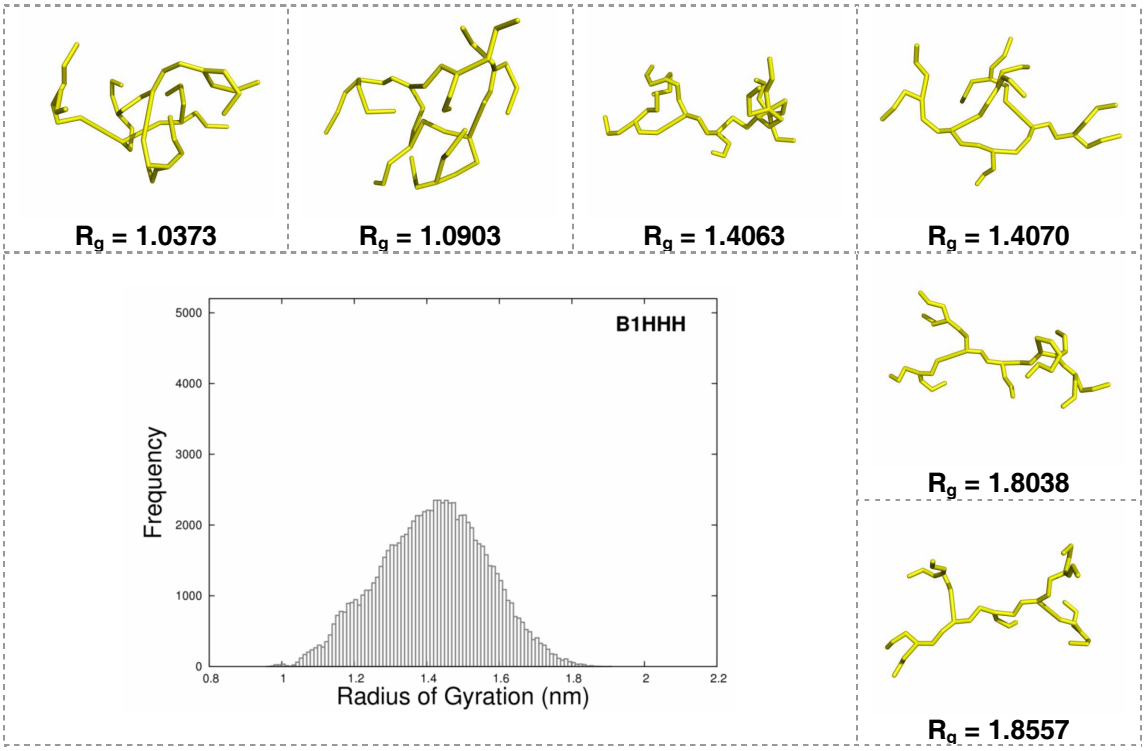


Figure 3.9 – Radius of gyration histogram for B1HHH.

In the center of the figure we present an histogram constructed with all the R_g values collected from the equilibrated trajectories. Additionally we present atomic configurations chosen as examples of the populations within certain bins (see the text for further details).

The analysis and main conclusions to be redrawn from these figures are almost identical to the ones for B1. In fact the shape of the histogram is similar (almost a Gaussian distribution) and no single R_g value stands from the rest.

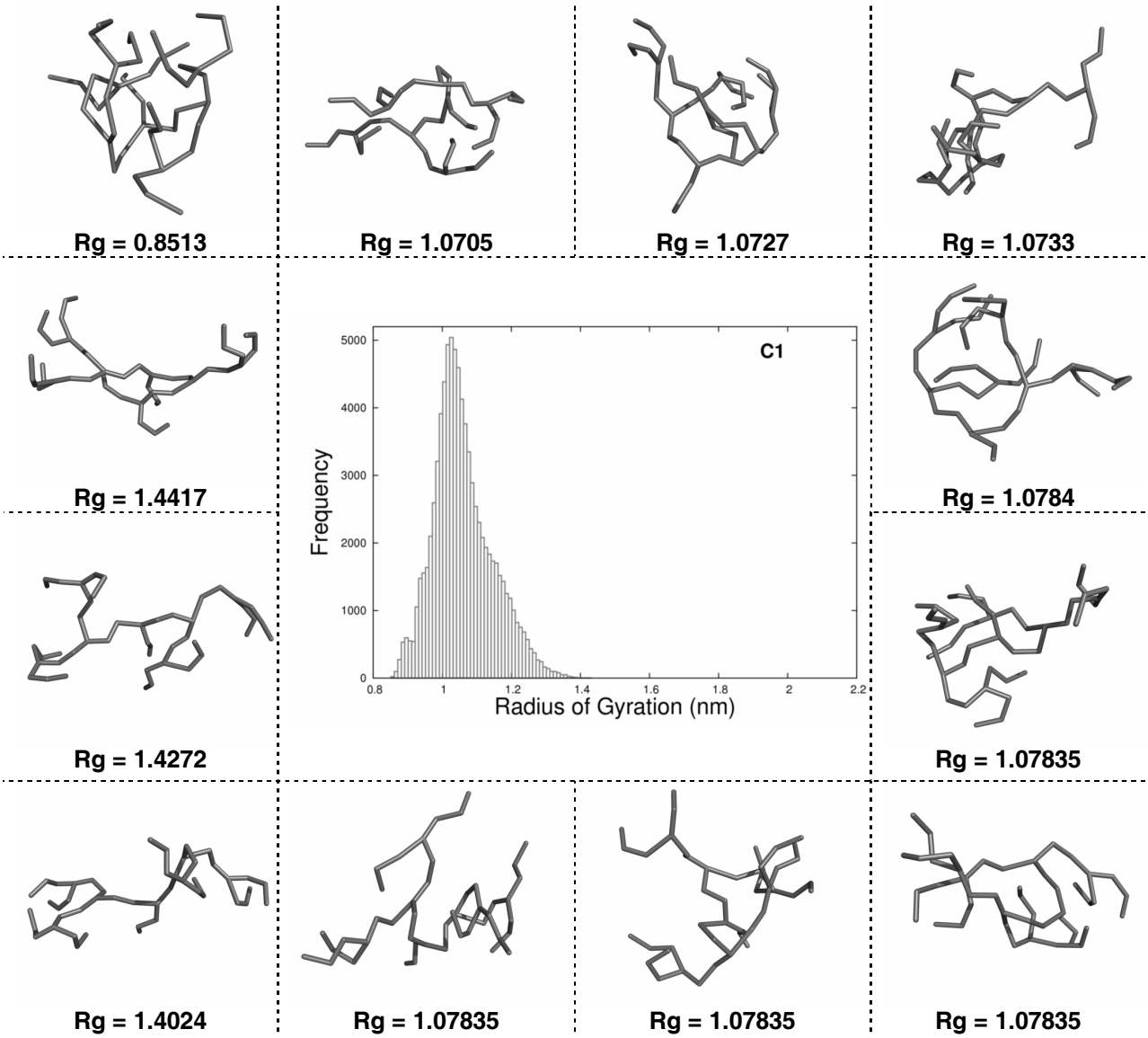


Figure 3.10 – Radius of gyration histogram for C1.

In the center of the figure we present an histogram constructed with all the Rg values collected from the equilibrated trajectories. Additionally we present atomic configurations chosen as examples of the populations within certain bins (see the text for further details).

Comparing with other peptide dendrimers the R_g values for C1 vary in a much inferior range (from 0.8 nm to 1.4 nm) depicting a smaller region, from 0.8 to 0.9 nm, and a bigger region between 1 and 1.1 nm.

Globally the shape of the C1 histogram is different from all the others presented before; with an obvious smaller range of preferred R_g values (from 1.3 to 1.7 nm). Even the smaller values for C1 are much smaller than the ones obtained for the other

dendrimers. This means that the combination of amino acid residues used to construct C1 promotes a more compact structure (by comparison with the other dendrimers under study).

Once more we showed pictures displaying the dendritic conformation of different replicates at different times of the simulation but that share the same R_g value. We verified that the most recurrent R_g values correspond to conformationally heterogeneous structures, sharing a common compactness but a very specific atomic configuration.

3.2.3.2. Comparison of R_g Histograms

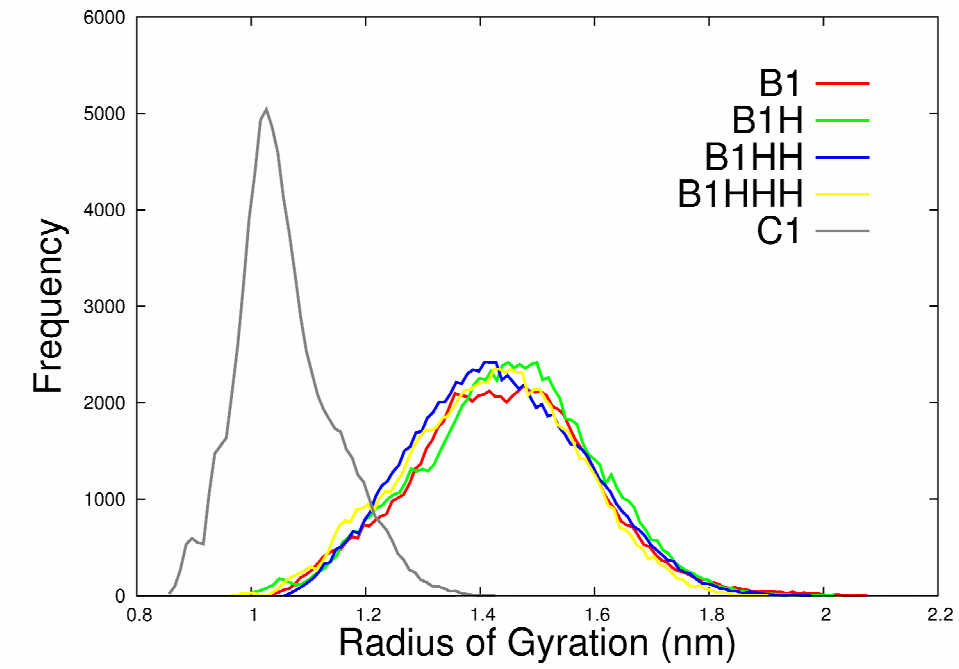


Figure 3.11 – Radius of gyration histograms for B1, B1H, B1HH, B1HHH and C1.

The R_g histograms presented in the previous figures (Figure 3.6, 3.7, 3.8, 3.9 and 3.10) are displayed in the plot allowing a better comparison. The lines corresponding to each dendrimer are identified using a color code, as is evidenced in the figure (see the text for further details).

The first conclusion from analyzing all histograms of R_g collectively is that they all present a "central" set of R_g values that are the most populated ones (see Figure 3.11). Nevertheless, small sections of the populations present structures that account for levels of compactness completely opposed to the most frequent ones. Although the relative percentage of these values is small, it evidences the expected conformational flexibility that characterizes peptide dendrimers (and dendritic structures in general)[31].

Moreover, comparing the histograms in Figure 3.11, we observe that all B1-family dendrimers present R_g values almost in the same range.

A noteworthy peak in the histogram is the one for B1H at 1.05 nm, where there seems to exist a higher set of low R_g values conformations then in any other B1-family dendrimer.

Furthermore, the C1 dendrimer presents R_g values in a smaller range than the others and with inferior R_g values, being a much more compact structure (as can also be concluded from the pictures surrounding the respective histogram, and by Figure 3.11).

An important conclusion can be assessed for B1H and C1, where we present images of conformations with the exact same values of R_g but with markedly different branch distribution and orientation (see Figure 3.7 and Figure 3.10). The images suggest that the histogram analysis performed allows for a characterization of the global compactness trends in different dendrimers (more "closed" or more "open") but is not a satisfactory conformational discriminant, because different topologies are attributed to the same bins of the histogram (i.e., the same values, different geometries).

As a conclusion it must be emphasized that our results point to the existence of a significant number of non-spherical conformations of dendrimers. This fact is evidenced by the average R_g values, calculated using all equilibrated trajectories. A conclusion that contradicts the ones presented by other authors [Javor 2009].

[31] See Chapter 1, section 1.9., on the impossibility of obtaining a crystallographic structure of peptide dendrimers.

Table 3.4 – Average radius of gyration values for the entire set of equilibrated trajectories.

Dendrimer	$\langle R_g \rangle$
B1	1.5415
B1H	1.4899
B1HH	1.5017
B1HHH	1.4188
C1	1.1220

Next we present a series of histograms drawn using other properties of the system. In this manner we expect to characterize the most frequent states of the system.

Some snapshots are presented for the B1 and C1 dendrimers to illustrate the relation between the discussed property and the dendrimer conformations.

3.2.3.3. Total number of hydrogen bonds

The total number of snapshots considered to plot the different histograms is the same employed to the R_g ones.

The analysis of the R_g values distribution suggests that interactions between branches in the dendrimer may play a key role in the overall conformation. To understand the variables involved in the formation of the structures depicting more ordered states, we next analyze the *hydrogen bond* (HB) interactions.

The first case we present, concerns the total number of intramolecular hydrogen bonds in each structure. The determination of these values was achieved using the GROMACS associated tool, *g_hbond*.

We defined HB using a geometrical criterion between potential donor (HBd) and acceptor (HBa) atoms. The OH and NH groups are regarded as potential donors, while O and N atoms are considered as potential acceptors. The total number of potential donors and acceptors for each dendrimer is represented in Table 3.5.

Table 3.5 – Number of potential hydrogen bond donors[a] (OH and NH groups), and potential hydrogen bond acceptors[b] (O and N atoms).

Dendrimer	HBd[a]	Hba[b]
B1	54	126
B1H	55	128
B1HH	56	130
B1HHH	57	132
C1	63	119

After determining if an atom is potentially involved in a hydrogen interaction we applied decisions based on cutoffs.

An HB is considered to exist if the HBd-HBa distance is equal or inferior to 0.35 nm,[32] and the angle formed by the hydrogen atom (H) and HBd-HBa (the H-HBd-HBa angle) is equal or inferior to 30º.

[32] The value of 0.35 nm for maximum HB distance is chosen in accord with the first minimum on the radial distribution function (RDF) for SPC-water.

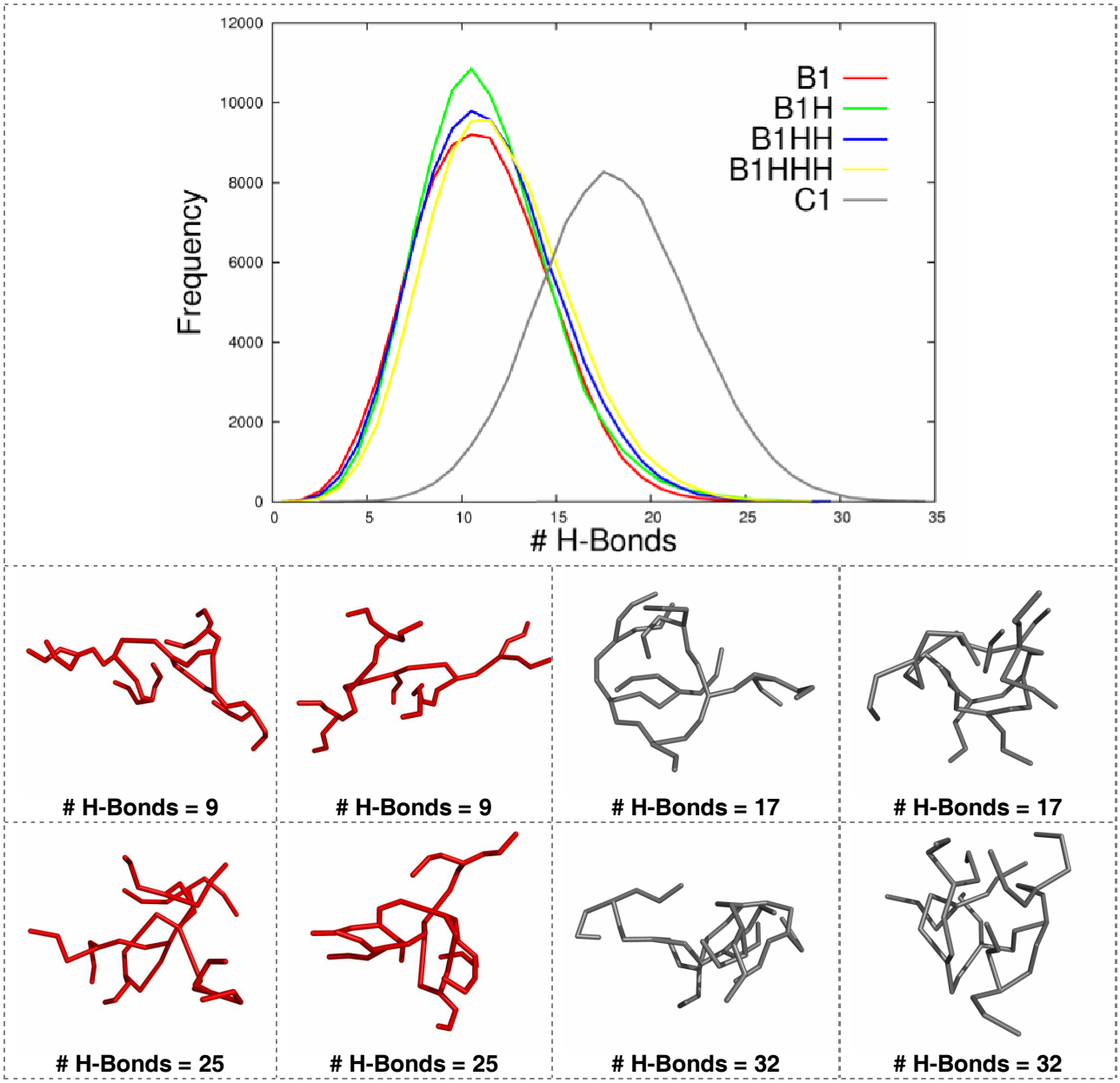


Figure 3.12 – Total number of hydrogen bonds histograms.

Histograms for the total number of hydrogen bonds (#H-Bonds) presented in every conformation in the ensembles considered for each dendrimer. Figures of some conformations corresponding to the most representative bins and to the ones with more HBs, are presented for B1 and C1 (see the text for further details). The conformations of B1 are represented in red, whereas the ones of C1 colored in gray.

From the B1 histogram we can se a data distribution resembling a Gaussian one, with an mode value of 11 HBs, representing approximately a fifth of the total potential HB donors. Furthermore, from the images we verify that, as expected,

structures with less HBs present themselves more loose than the ones with a higher number of HBs. Accordingly, we verify that the more "ordered" and compact structures have approximately 25 HBs and represent a very small section of the entire available conformations.

Observing the histograms for B1, B1H, B1HH and B1HHH we noted that the ideas stated for B1 are also valid for the rest of the B1-family. The most frequent value is 11 HBs or near this value; all histograms have similar shapes and ranges (they all vary between zero and twenty five).

Contrary to the homogeneity within the B1-family, the C1 dendrimer has values for the total number of HB that range from 3 to 35, with a mode value of approximately 17 HBs, and where the structure branches seem to interact often.

As can be ascertained from the adjacent images, different structures of C1 can present the same total number of HBs, although they correspond to interactions between different atoms.

When comparing C1 and B1 we conclude that C1 has a greater number of HBs, that probably account for part of the greater compactness evidenced in the R_g histograms.

In effect, no clear trend or conformational preference can be detected.

3.2.3.4. Solvent Accessible Surface Area

The total solvent accessible surface area (SASA) of a molecule is (as the name suggests) the total area exposed to interaction with solvent molecules. This parameter/property was also determined for each dendrimer through the use of a GROMACS tool (*g_sas*) that uses a spherical probe (with 0.14 nm of radius corresponding to a water probe) to scan the entire molecule surface. This parameter is deeply related to the compactness of a structure (and therefore with the radius of gyration) and with the molecular volume.

To construct the corresponding histogram we have used bins with 0.5 nm^2.

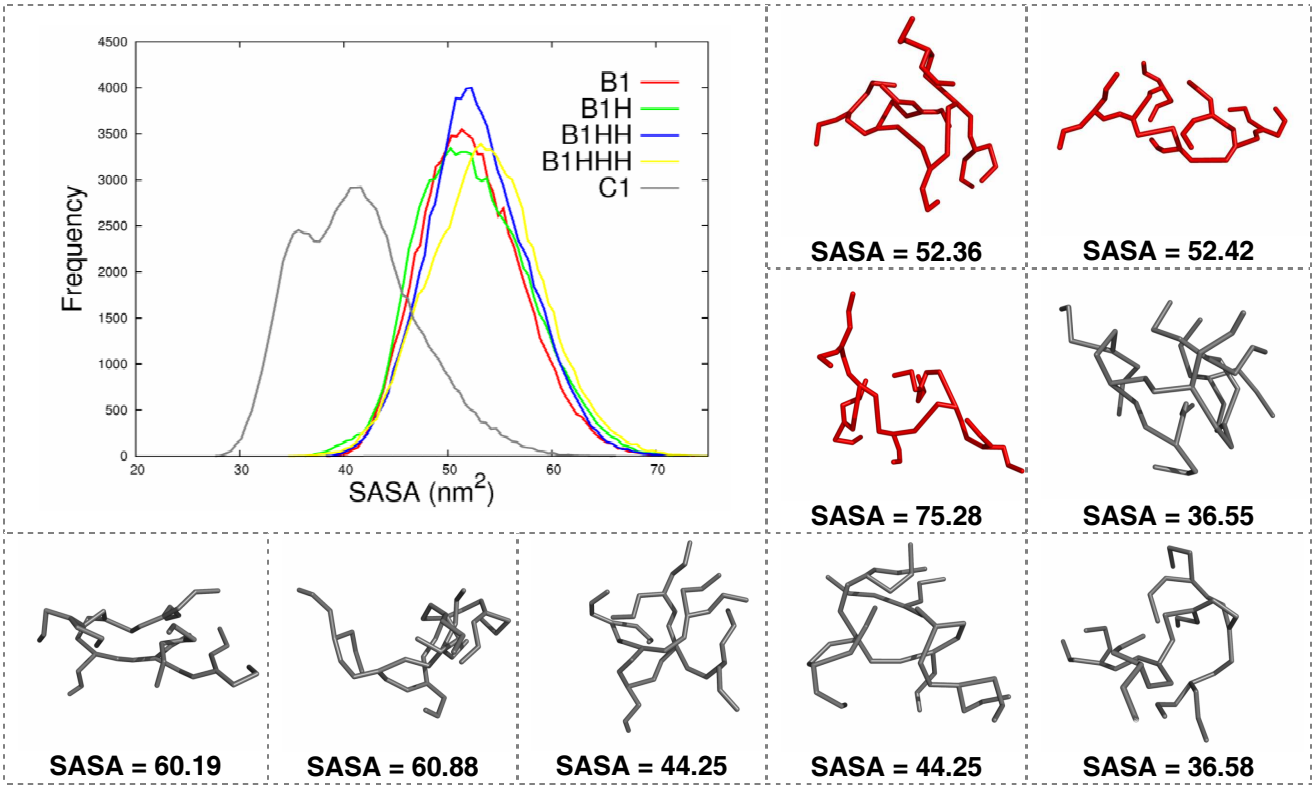


Figure 3.13 – Solvent accessible surface area histograms.

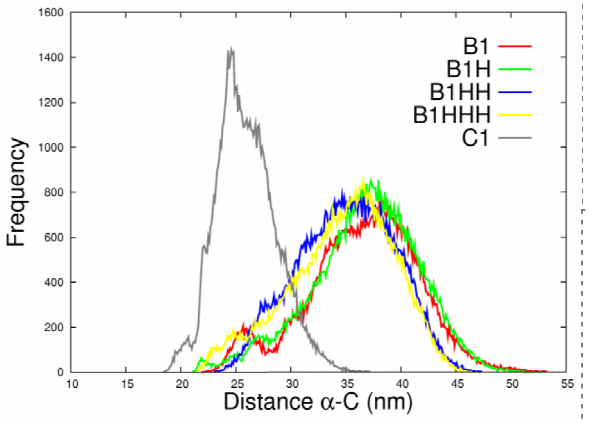
Histograms for the solvent accessible surface area (SASA) of every conformation in the ensembles considered for each dendrimer. Figures of some randomly selected conformations, corresponding the most representative bins and to the ones with the higher or minor values of SASA, are presented for B1 and C1 (see the text for further details). The conformations of B1 are represented in red, and C1 in gray.

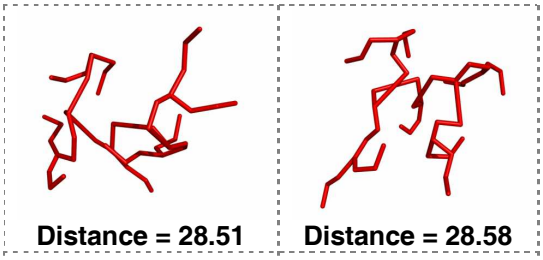
We find significant water penetration inside the dendrimer molecules. The results obtained for B1, B1H, B1HH and B1HHH are in accordance with the ones obtained using other properties to perform the same analysis (R_g and HB). The SASA values range from approximately 40 to 70 nm^2 and the most frequent values are "placed" near 50-55 nm^2.

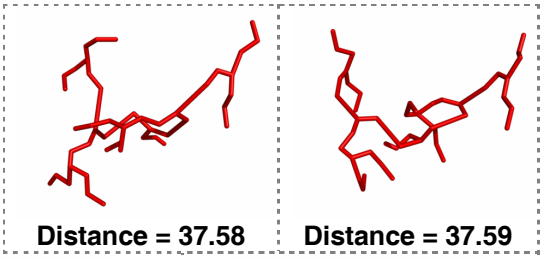
However, for C1 a two-state behavior seems to appear once more, with the histogram depicting two distinct lumps. These lumps are placed very close in the histogram and the conformations that constitute each of them are very heterogeneous. With the first lump (ranging from 34-37 nm^2) presenting greater intra-branch contacts and a more compact structure than the structures composing the 43-47 nm^2 bins.

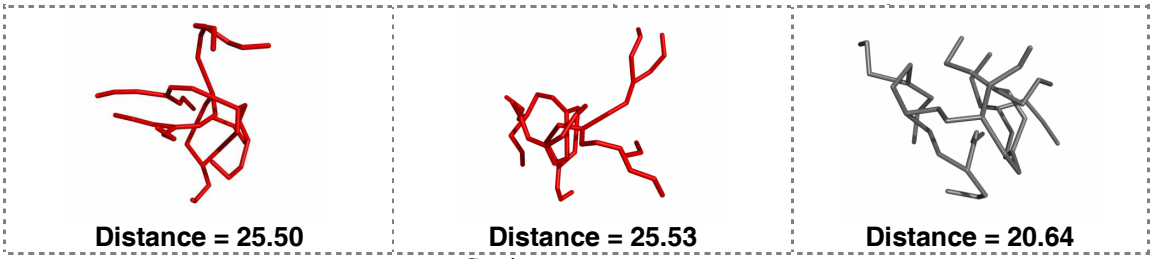
3.2.3.5. Sum of the distance among the branching units alpha carbons

We have defined the sum of the distance between the alpha carbons of each branching residue (in a number that is always equal to seven whether they are LYR or DAP residues) as a measure for the branching distribution within each dendrimer at a specific time (or in a specific conformation). The corresponding histogram is presented next using bins with a 0.1nm size.









Continues on next page

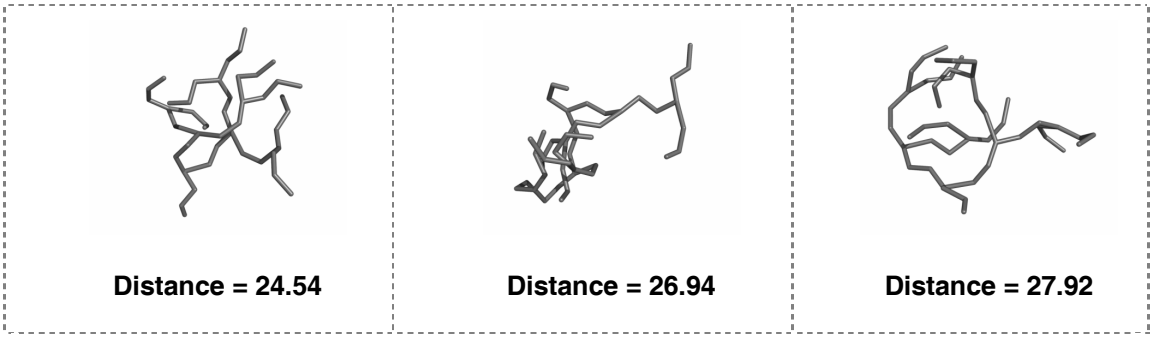


Figure 3.14 – Sum of the distance among the branching units alpha carbons histograms.

Histograms for the sum of the distance among the branching units alpha carbons (Distance α-C) of every conformation in the ensembles considered for each dendrimer. Figures of some randomly selected conformations, corresponding the most representative bins and to the ones with the minor values of Distance α-C , are presented for B1 and C1 (see the text for further details). The conformations of B1 are represented in red, and C1 in gray.

Observing the resulting plots one can not avoid comparing them to the histograms of R_g, in fact, although they display significantly more detail (due to the small size of the bins) the overall shapes drawn resemble the ones of R_g with similar tendencies.

Furthermore, if we superimpose the R_g and the sum of the branching units alpha carbons distance values of each replicate, we verify that the obtained variations are almost identical (data not shown).

The only noticeable exception, where further information is obtained from these histograms, are B1 and B1H, where two distinct lumps are observed for different ranges of the Distance α-C. Still, the conformational composition of each of these lumps is too heterogeneous to draw any further conclusions.

3.2.3.6. Distance between the two farthest atoms

As a complement to the information and histograms previously presented, a histogram constructed using as parameters the distance between the two most far apart

atoms in the dendrimer at each time, is displayed next. The values were calculated using *g_dist* and the bin size was 0.1 nm.

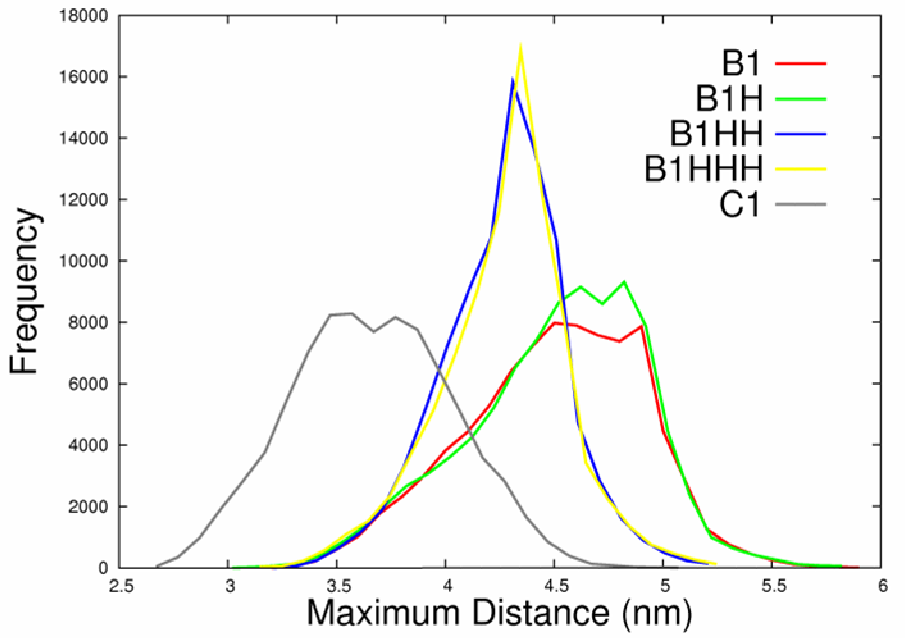


Figure 3.15 – Distance between the two farthest atoms in the dendrimers.

Histograms for the distance between the two most far apart atoms (Maximum Distance) of every conformation in the ensembles considered for each dendrimer (see the text for further details).

We consider this measure as a descriptor of the maximum linear distance occupied by each conformation.

As expected the B1-family dendrimers present maximum distances in similar intervals (3-5.5 nm), and in a wider range of values than C1. Meaning that in its most stretched conformations C1 is smaller than any of the other dendrimers most stretched conformations.

Moreover, dendrimers presenting histine residues at a X_4 position (see Figure 1.4) demonstrate mode values that are inferior and more frequent (almost the double) then the ones characteristic of dendrimers with Amb (B1 and B1H) at that position.

As there are only some small differences in key residues between the B1-family dendrimers, it comes without surprise that the range of maximum distance values accessible to each dendrimer depends on its amino acid composition (because they all

present the same branching pattern and branching units – See Table 1.2). For that reason we can hypothesize that the existence of a histidine residue at position X_4 (see Figure 1.4, Chapter 1) promotes a shortening of the "main chains" of both the branches that comprise this position. This might happen because the "atomic contribution" of His is smaller than the atomic contribution on an Amb residue[33]. And as such, the Amb "size contribution" allows for a greater stretching of the entire dendrimer and a greater structural flexibility.

While His accounts only for a C_α, a C and a N atoms, the Amb residue incorporates into the main chain almost its entire atomic structure, including an aromatic group (see Figure 2.3), this allows both branches that include X_4 (see Figure 1.4) to have more "stretching potential" and the possibility to undergo less compact states when using Amb (Amb is in fact a better spacer molecule, with reflection in the properties exhibit by the corresponding peptide dendrimers). According to this idea, and as can be ascertained from the B1 and B1H histograms, the single exchange of a Cys by a His residue in X_2 (see Figure 1.4) has little influence in the maximum distance.

3.2.3.7. Summary of histogram results

In every analysis of MD results (and normally in any analysis of results), one starts by employing the methodologies and tools that are more readily accessible and easy to implement according to the objects of study. We constituted no exception to this rule, ergo, we started our analysis by plotting different properties as a function of time to access the global quality of our data and the intrinsic variation incorporated on it. Afterwards, we plotted the values of different properties in function of each others to try to find specific correlations among them (see section 3.1.). Finally (up to this

[33] By atomic contribution we mean the number of atoms belonging to the amino acid residue that are incorporated in the corresponding main chain. Normally a proteinogenic amino acid "contributes" with three heavy-atoms to the main chain, two carbons and a nitrogen one.

point) we have performed a detailed histogram analysis of our data in an attempt to identify differences in the probability on which the different conformations occur.

Unfortunately, and although these analysis were useful and important to account for the overall features of the different conformations, they do not reflect/discriminate in a satisfactory manner the conformational variability characterizing the dendrimers. Still, we must emphasize that this approach has allowed us to quantitatively distinguish, and characterize, between dendrimers with major differences in their amino acid residues composition, but that share the same dendritic architecture.

At this point, we can conjecture that the C1 dendrimer seems to present at least two types of distinct conformational preferences. However, the approach employed to suggest this hypothesis does not satisfactorily discriminates the conformational differences among the structures.

With this in mind, we must consider that there may be a better way to analyze these trajectories, and that those other approaches might reveal different trends and yield a better discrimination of the "conformational space". So far, all evidences point to a monotonic behavior by part of the B1-family dendrimers (if we think on the presented results in terms of energy curves and surfaces) with a kind of "downhill folding". Based solely on the so far provided results we must consider that some dendrimers exhibit a great conformational flexibility, where small energy variations are responsible for an easy inter-conversion among different structures. The path for these inter-conversions remains unstudied here, and the existence of kinetic intermediates is not discussed, nor is it a mandatory waypoint in this thesis. Even so, we must consider that some dendrimers (in particular C1) evidence a two-state behavior.

We also found evidences suggesting that no particular shape/form is exclusive of these dendrimers and, although they might undergo more compact shapes, they do not resemble spheres (as some authors have defended [Javor 2009]).

The first discussions on dendrimers and the commonly employed schematic descriptions rendered the idea that, the dendrimer branches might be almost equitably disposed around the molecules core, with the probable existence of a central cavity and the terminal groups forming the peripheral surface. However, we found

consubstantiated evidence on the existence of significant "back-folding" of the end groups towards the dendrimers core.

3.2.4. Phi and Psi torsional angle distribution

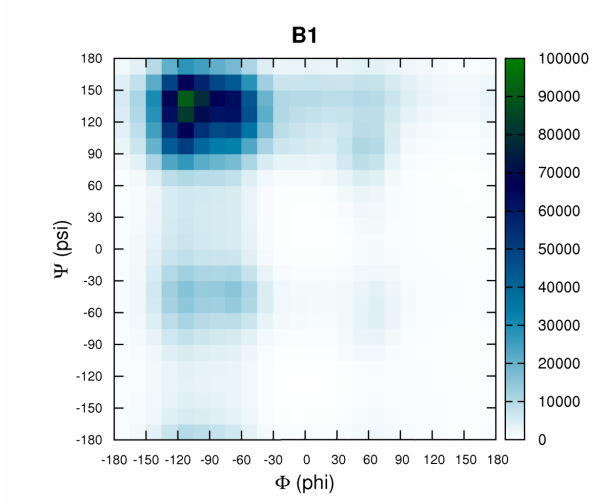
Ramachandran plots[34] [Cantor 1980; Leach 2001] are a good way to visualize the possible backbone conformations associated with the dihedral torsion angles (*phi* and *psi*) of molecules containing peptide bonds. Drawing these plots is a common approach when dealing with polypeptides, and as we are dealing with peptide dendrimers, useful information can be obtained in this way.

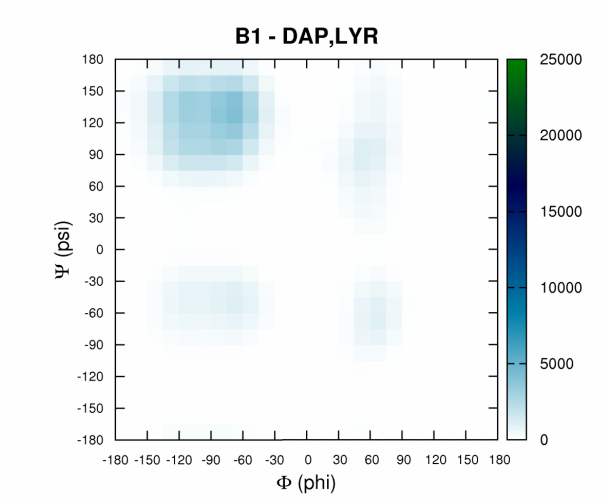
The main question we intend to answer is whether particular values for the dihedral angles arose due to the distinctive architecture shared by dendritic molecules; placing special emphasis on the *phi* and *psi* values adopted by the different branching residues.

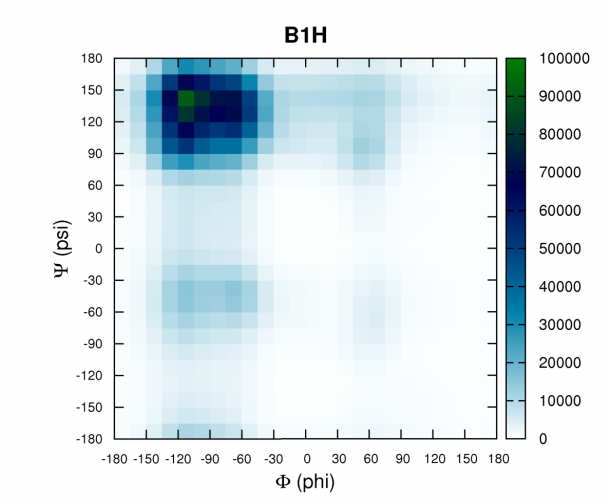
Furthermore, we mean to discuss the existence of differences between dendrimers containing miscellaneous residues as branching units, specifically Lyr and Dap (B1, B1H, B1HH and B1HHH), or the same residue as branching units, such as C1 with Dap.

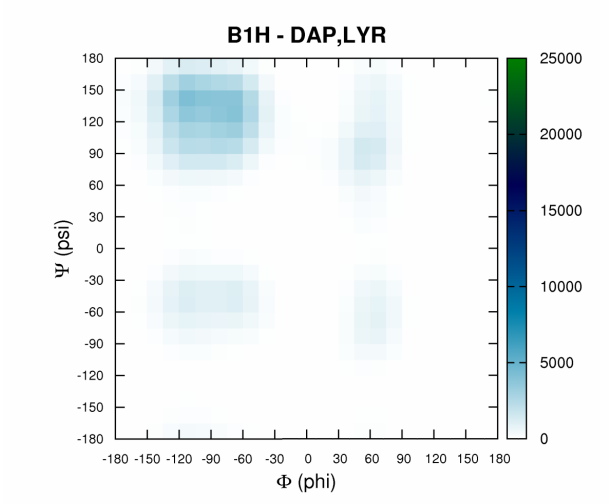
We have collected the values for the dihedral angles of every snapshot of the equilibrium concatenated trajectories using some in-house scripts and the GROMACS *g_rama* program. The values are represented in Figure 3.16 on the plots to the left. On the right plots we have represented the dihedral angles for the respective branching residues. The type of branching units present in each peptide dendrimer is mentioned in the corresponding title.

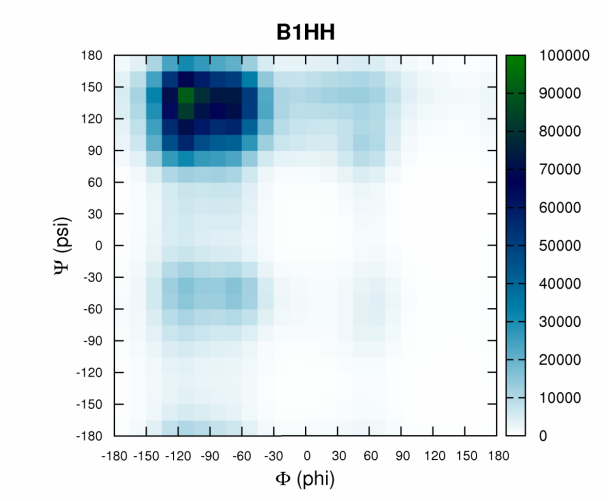
[34] Due to steric hindrance, the main chain of a polypeptide usually assumes preferred, energetically favorable conformations. For each residue, these conformations can be characterized by the value of two torsion angles, phi and psi. The phi angle of residue i is defined by the torsion Ci-1-Ni-Cαi-Ci, and psi by the torsion Ni-Cαi-Ci-Ni+1.The distribution of phi and psi is usually called the Ramachandran plot [Cantor 1980; Kleywegt 1996].

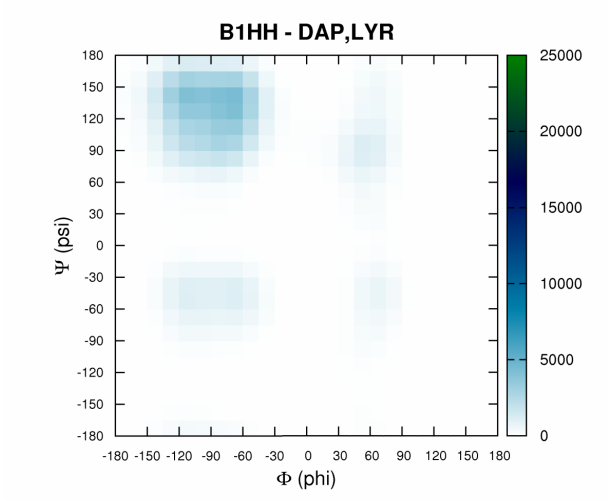












Continues on next page

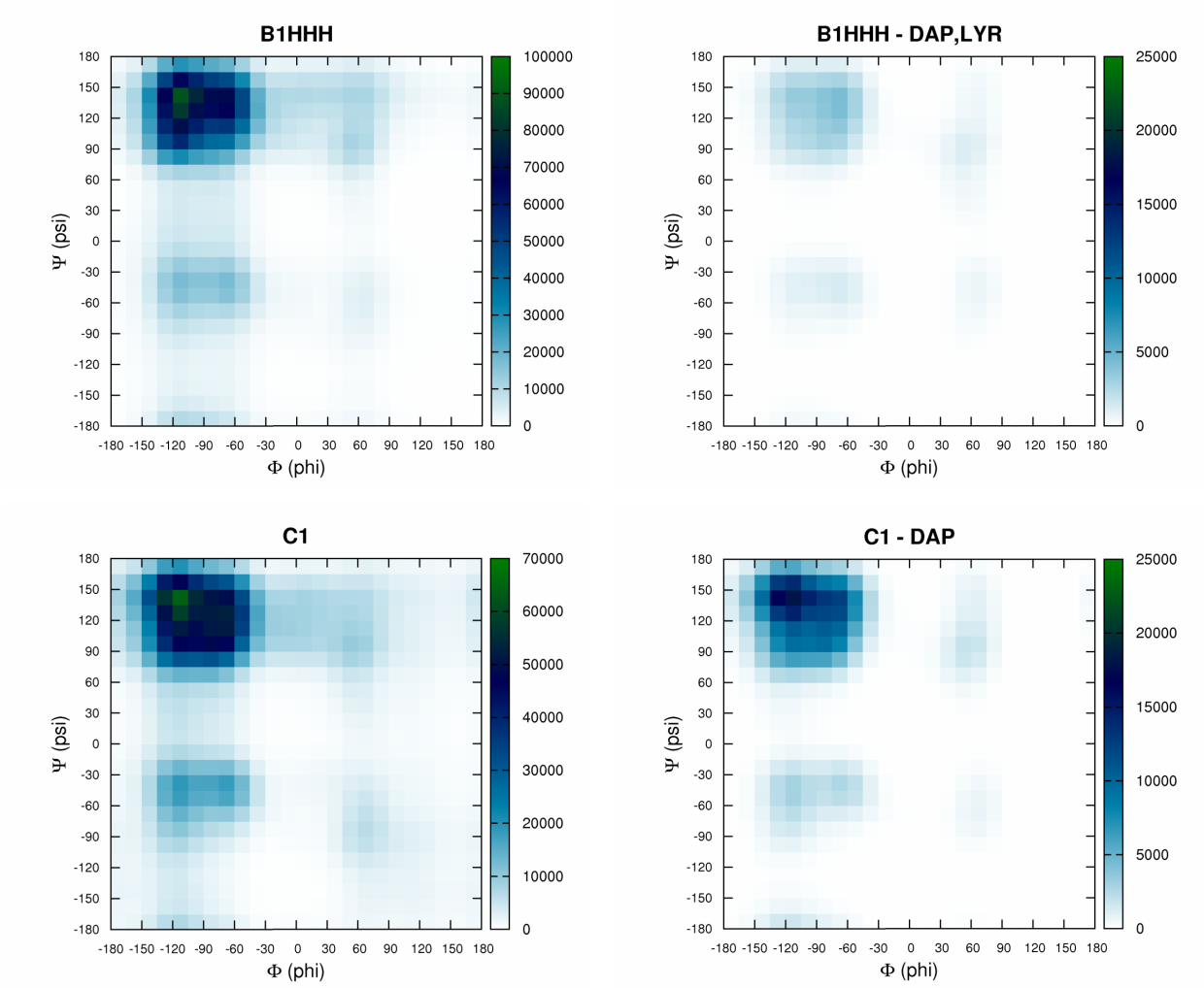


Figure 3.16 – Ramachandran plots for B1, B1H, B1HH, B1HHH and C1 and the respective branching residues.

The phi and psi dihedral angle values are in degrees. The plots on the left represent all the torsion angles of each dendrimers; the ones on the right account only for the torsion angles of the branching residues, either a combinations of Lyr and Dap (B1, B1H, B1HH and B1HHH), or just Dap (C1). See the text for further details.

The primary conclusion obtained from the plots presented is that, the possible combinations of *phi* and *psi* angle values favored by these peptide dendrimers values are in accordance with the ones exhibit by most polypeptides [Lovell 2003; Hovmoller 2002; Cantor 1980].

The typical areas permitted for protein dihedral angles are all present in peptide dendrimers, with greater predominance of torsion angles with negative *phi* and positive *psi* values.

The plots collecting all *phi-psi* values are similar for all dendrimers, regardless of the specific residues composing them. The type of residues selected as branching units seems to have little influence on the allowed dihedral angle values, as can be ascertained comparing the plots for the dihedral angles of all branching residues.

When comparing the phi and psi values obtained for dendrimers using both Lyr and Dap as branching unit, or dendrimers using just Dap we observe that the areas of the plot occupied are the same. Therefore, when analyzing simply the torsion values to the dendrimers that use either Lyr or Dap as branching residues, the branching residues does not seems to be a major conditioning factor in the shape of the structure.

Given the fact that scientific literature information on peptide dendrimers *phi* and *psi* angles does not exist, one can conclude that the values presented here are reliable (based on the validation and equilibration considerations aforementioned).

3.2.5. Contact Maps

Another useful approach to discuss the interactions among different residues in a molecule is to use residue *contact maps*. These maps express the smallest distance between every pair of residues in the molecule through a two-dimensional matrix [Vassura 2008; van der Spoel 1998(b)].

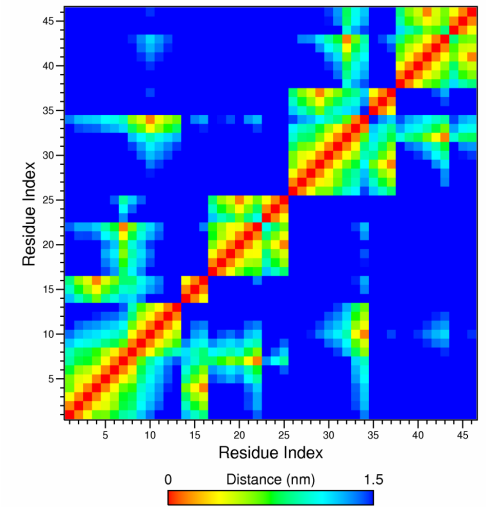
Similarly to other authors, we defined the smallest distance between two residues as the smallest distance between any pair of atoms. The distance matrix is thus symmetric by definition [van der Spoel 1998(b)]. This matrix contains in its entries the Euclidean distance between each pair of residues. If the total number of dendrimer residues is represented as N, we need N^2 elements. Since the matrix is symmetric the effective number of needed elements is only $N(N\text{-}1)/2$.

We have computed these matrices for each of our peptide dendrimers using the *g_mdmat* program (GROMACS) on the equilibrated and concatenated trajectories.

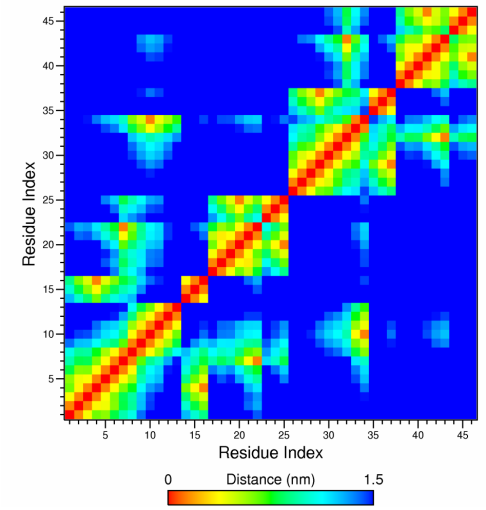
3. Results and Discussion

For the sake of computational tractability we have only considered conformations collected over intervals of 200 ps from these trajectories. The total number of snapshots/conformations considered is equal to the total number of equilibrated snapshots dividing by two hundred and adding one.

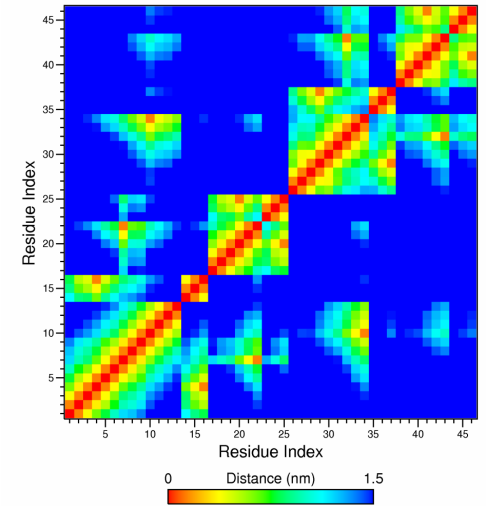
B1



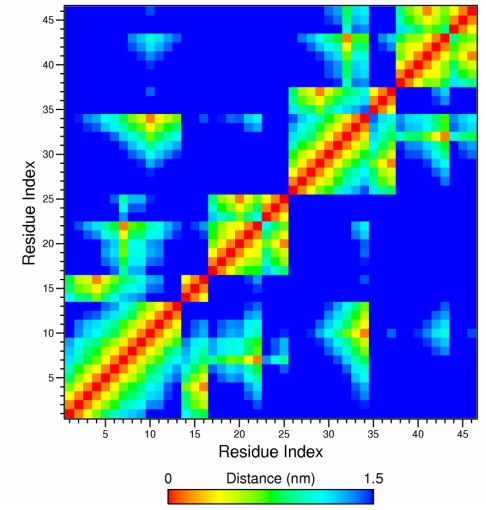
B1H



B1HH



B1HHH



Continues on next page

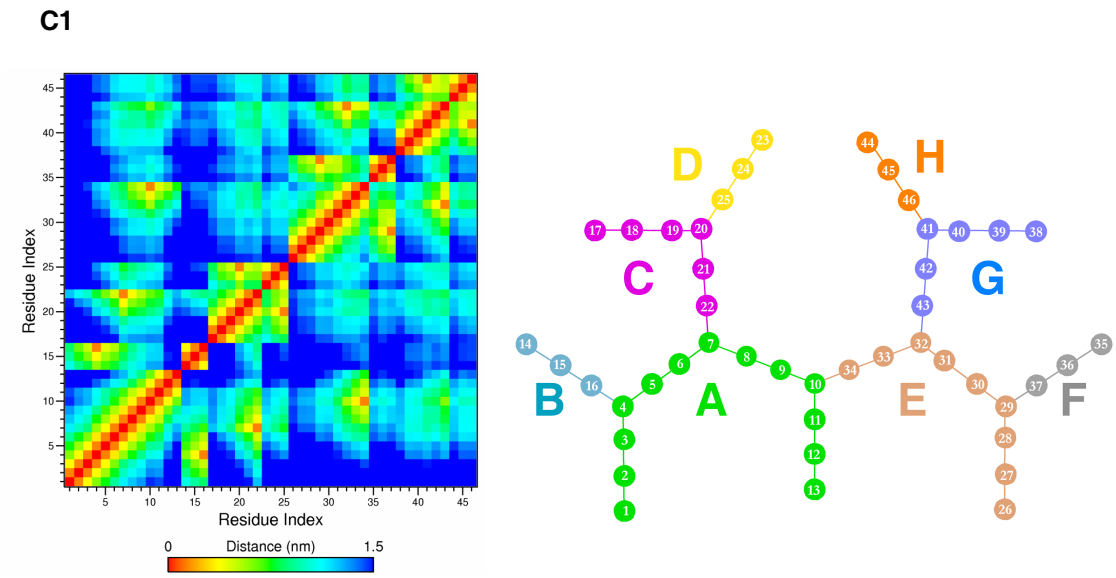


Figure 3.17 – Residue-residue distance matrices for B1, B1H, B1HH, B1HHH and C1.

To construct these matrices we have used: 4251 (B1), 4526 (B1H and B1HH), 4401 (B1HHH) and 4176 (C1) frames. A schematic representation with numbering of the different residues and the associated chain names is presented at bottom (on the right). The distances between the residues can be evaluated through the color scales depicted in the figures.

To conveniently analyze these matrices the reader must have in attention Figure 2.6 presented at Chapter 2.

Although these maps may seem, at a first glance, to contain little information, they carry, in fact, discriminative details on the possibility of interaction among the different residues comprised in a dendrimer. Understanding these maps means understanding the allowed and disallowed proximities within each pair of residues in a peptide dendrimer. From the matrices, we once more verify the similarity between all the B1-family dendrimers, with the contact maps for each of the family members being almost identical.

Contact maps methodology was developed to study proteins, which are basically linear compositions of consecutive amino acid residues. When applying this concept to dendrimers we must keep in mind that these are branched molecules and that, although we have defined a convention for the numbering of residues based on linear peptide chains, some chains will encompass more residues than others (see in the

figure). This is the problem associated with defining and cataloging branched symmetrical (or quasi-symmetrical) chemical structures. And it is impossible to circumvent it because almost all tools/programs that resort to mathematics need a way to sequentially differentiate the input. In the case of dendrimers the ideal situation would involve a chain nomenclature able do discriminate between residues assigned to the same relative position while considering them as numerical equivalents.

In the particular analysis of contact maps this classification problem has little importance, but it is responsible for the occurrence of special patterns where the distance between residues that are connected and/or adjacent, although small, might not appear at the diagonal of the matrix. This is the case of the distances between residues 9 to 13 (from chain A) and residues 31 to 34 (chain E), that are very close in reality and bonded through residues 10 and 34.

The best way to view dendrimers contact maps is by analyzing not only the pair of residues interacting but also its relative position within the dendritic geometry (core, branching or periphery).

The contact maps for the B1 dendrimer show a filled diagonal, indicating the interactions among consecutive residues, in fact we can distinguish nine data agglomerations near the matrix diagonal, each one corresponding to a dendrimer chain and one that accounts for the residues 8 to 13 from chain A, including the core residues.

These diagonal agglomerations are surrounded by small distance values as a consequence of interactions among adjacent peripheral residues at the end of each chain, for example the visible interaction among chains D and C.

Observable interactions within the considered distance occur between multiple chains, with the first branching unit (counting from the core, position 10 in the scheme) clearly separating the residues that interact, meaning that the residues of chains A, B, C and D are able to interact among them but can not interact with the residues form chains E, F, G and H (that also interact among them). This means that we can denote the existence of “two main branches” each composed by four chains, the exception are the residues from 10 to 13 that constitute the central core and that by a matter of practical classification where included in chain A. The core of B1 interacts not only with the residues that are bounded and adjacent to it but also with some

residues from peripheral groups, namely chain H. This implies that there exists a high flexibility around the first branching residue and that the dendrimer arms defined at that point always present greater interaction among residues from that arm than inter-arms residue interactions.

Still we must emphasize that in B1 the smallest distances between non-subsequent residues is mainly over 1.5 nm excluding the existence of a tight interwoven structure embracing simultaneously all defined chains.

The contact map for B1H is, in terms of the residues that contact with one another, identical to B1, and as such the conclusions obtained for B1 are also valid for B1H. A relevant consideration when comparing the contact maps of B1 and B1H with the ones form B1HH and B1HHH is that a significant difference occurs between residues 42-43 in chain G, and residues 9 to 13 from chain A. The distance between these residues is significantly smaller in dendrimers containing histidine at positions 8 and 33. This two residue difference is also responsible for the disappearance of an "agglomeration" in the diagonal of the matrix. While in B1 and B1H we observe the existence of nine clear groups in the diagonal, in B1HH and B1HHH we can only distinguish eight groups, with the residues from 1 to 13 constituting a "distance-homogeneous" group.

This observation means that the histidine residues at positions 8 and 33 are able to come closer to their surrounding residues than Amb residues at the same position. This probably happens due to the long side chain present in histidine which is absent in Amb.

The C1 residues compose a more complex contact map filled with inter-residues interactions. Similarly to B1HH and B1HHH, in C1 we also observe a filled diagonal, indicating the interactions among consecutive residues. We can distinguish eight data agglomerations near the matrix diagonal, each corresponding to a dendrimer chain.

However, in C1 we verify the existence of interactions among branches that are, at least "schematically", distant. We observe small values for the distances among the core residues and chains E (which is bounded to the core), C, D and G (that are not bounded or adjacent to the core). Furthermore, if we define "two main dendrimer arms" using the bifurcation in position 10 (as done for B1) we verify small distance values for pairs of residues composed by residues of different arms (e.g. residues 19

to 22 with residues 41 to 43). In fact, the main contacts between these two arms occur not between peripheral residues but between the more "central residues" (positions 5 to 9, 20 to 22, 32 to 34 and 41 to 43).

Despite these distant branch contacts in C1 contact map we also observe small values for the distances between third generation spacer residues that are bounded to the same branching residue (as would be expected due to their "forced" atomic proximity).

Based on the aforementioned conclusions we can consider that C1 has the potential to present conformations with a great number of residues close to each other, and a number of intra-dendrimer contacts. This idea is probably related with the maximum total number of hydrogen bonds allowed for the C1 dendrimer (see Section 3.2.3.3.).

3.2.6. Algotithmic Clustering

Molecular dynamics simulations provide valuable insights into the structure and interactions of macromolecules. This knowledge is obtained through the study of different molecular properties. However, even with the elucidation of these properties, often the inherent relationships among the molecular configurations are hidden in the complexity of the data [Shao 2007; Everitt 2001].

We have generated a large amount of conformational data and it is desirable that we identify a smaller, set of representative conformations.

A useful, and widely [Shao 2007] used, approach to expose some of these correlations is to cluster (group classification) molecular geometries based on their degree of similarity (measured by an appropriate metric) [Everitt 2001]. Clustering is the unsupervised classification of patterns into groups based on their similarity. It allows (through the use of an appropriated algorithm) the partition of data points into a disjoint collection of sets named clusters. This approach can be applied to any data set regardless of the context as long as a pair distance measure function exists [Shao 2007; Everitt 2001]. Ideally, the molecular structures assigned to a cluster are closer among themselves than to structures from other clusters.

As the results obtained by clustering will generally vary as a function of the applied algorithm, and with the similarity measure cutoff chosen to consider structures as part of the same cluster, we can consider that applying cluster analysis to sets of data implies some assumptions and is, in fact, a somewhat heuristic and subjective approach.

The heuristic nature underlying clustering analysis conditions the validity and usefulness of the results. As a structural analysis tool clustering is only useful if it can "provide an unbiased mean of exposing significant relationships and differences in the underlying properties" [Shao 2007] and unfortunately this is not always possible. Moreover, applying different clustering algorithms to the same data will produce different results and there is no "universally better" clustering algorithm.

We have applied clustering methodologies to our structures using only a representative set of conformations because of computational tractability. The cluster analysis performed was done using 860 (B1),915 (B1H), 915 (B1HH), 890 (B1HHH) and 845 (C1) structures; what corresponds to select molecular configurations at 1 ns interval from the equilibrated trajectories of each replicate.

The clustering program employed is incorporated in the GROMACS package under the suggestive name of *g_cluster*. The *g_cluster* tool includes multiple clustering algorithms that employ the same (dis)similarity measure, namely the popular *root-mean-square deviation* (RMSD, see the definition further in the text; Section 3.2.7.). The use of RMSD produces clusters in which the molecules have a similar shape.

We have decided to use two clustering algorithms: Single Linkage [Everitt 2001] and Gromos [Daura 1999].

Single Linkage is a hierarchical agglomerative method that adds a structure to a cluster when its distance to any element of the cluster is less than a certain cutoff [van der Spoel 2005(b); Everitt 2001].

On the other hand, the Gromos method is a sort of divisive algorithm. It counts the total number of neighbors using the cutoff value, and afterwards the structures with the highest number of neighbors are taken as the center of a cluster, forming the cluster together with its nearest neighbors. The conformations that compose these

initial clusters are eliminated from the available pool of conformations and the procedure is repeated until all structures are accounted for [Daura 1999].

A major bottleneck in cluster analysis is deciding the adequate *cutoff* value to consider structures as neighbors; this problem is equivalent to choosing the ideal number of clusters describing the relations among the data.[35]

To assess the "correct" cluster count, we have used an analogy with the "elbow criterion" [Shao 2007]. This is a common *a posteriori* evaluation tool that chooses the appropriate number of clusters by noting where adding in additional clusters does not add sufficient new information. A representation suited to address this issue can be seen in Figure 3.18 where we plot the total number of clusters obtained while iterating over the RMSD cutoff value.

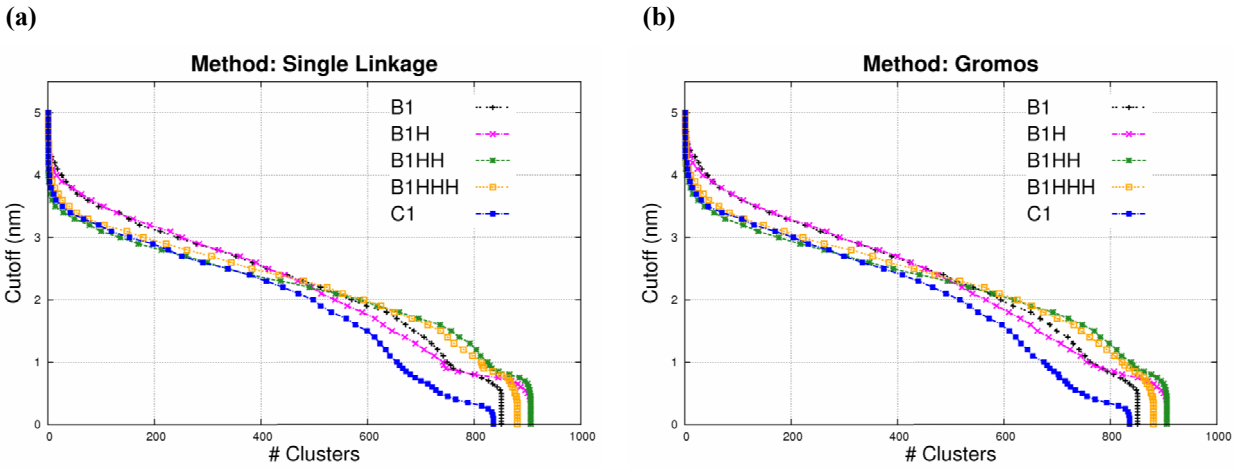


Figure 3.18 – Number of cluster obtained for different RMSD cutoffs using two different clustering algorithms, Single-linkage and GROMOS.

(a) Number of clusters (# Cluster) obtained while iterating over the RMSD cutoff value using the Single-linkage method for clustering the structures. **(b)** Number of clusters (# Cluster) obtained while iterating over the RMSD cutoff value using the GROMOS method for clustering the structures. See additional details in the text.

[35] This approach can be viewed as building the cluster tree (or *dendrogram*) for each dendrimer, and deciding where to cut it in order to obtain a significant number of clusters.

If the cluster algorithms applied are able to produce satisfactory results we should observe a vertical linear interval (a "stable region") illustrating the optimal cluster count, representing the cutoff values where changing the RMSD cutoff would not alter the total number of clusters.

As can be seem from the previous figure this does not happen, and accordingly we must consider that with the used algorithms (Single Linkage and Gromos) there are no evidences of meaningful clusters underlying our data. Ergo, either the molecular conformations do not present intrinsic geometrical correlations, or the clustering approaches used are not suited for revealing them. This results can be a consequence of various circumstances: the number of conformations considered for clustering may not be sufficient to represent all the conformational trends; the similarity measure employed might not reflect the "true" differences among the different conformations; the data used might be uniformly distributed over the considered space; among others [Everitt 2001].

In effect, in Figure 3.18 we observe that the only vertically linear areas of the curves correspond to count all the structures together as a single cluster, or each structure individually as a cluster, thus obtaining as many clusters as the number of input structures. This is, in terms of clustering analysis, meaningless.

The clustering methods used have proven not to be suited for the analysis of configuration preferences in peptide dendrimers. Therefore, we propose to base the "membership" to a particular conformational cluster on the topography of the estimated free energy surface; as was already done by other authors [Hamprecht 2001; Campos 2009; Becker 1997(a)] (see in the next section).

3.2.7. Root-Mean Square Deviation Analysis

The cluster analysis of the previous section is nothing more than a conformational analysis process to access structural preferences using RMSD similarities between all structures fitted on to a reference one. As mentioned before, the root-mean-square

deviation (RMSD) is a (dis)similarity measure used for quantitative comparison of pairs of structures [Becker 2001].

In Cartesian coordinates the RMSD between conformation *A* and conformation *B* of a given molecule is defined as the minimum of the function [Leach 2001; van der Spoel 2005(b)],

$$\mathrm{RMSD}(A,B) = \left[\frac{1}{\sum_{i=1}^{N} m_i} \sum_{i=1}^{N} m_i \left\| \mathbf{r}_i(A) - \mathbf{r}_i(B) \right\|^2 \right]^{\frac{1}{2}}, \qquad \text{Equation 3.6.}$$

where N is the number of atoms in the summation, i is an index over these atoms, and $\mathbf{r}_i(A)$ and $\mathbf{r}_i(B)$ are the positions of atom i in conformations A and B. The minimum value of Equation 3.6. is obtained by an optimal superposition of the two structures. The RMSD of a structure fitted into itself should have the value of zero.

Since the summation in Equation 3.6. may be on any subset of atoms, it can be finetuned to best suit the problem at hand. The summation may be over the whole molecule, but it is very common to calculate conformational distances based only on non-hydrogen heavy atoms or, in the case of proteins, even based on only the backbone Cα atoms [Becker 2001].

The RMSD values resulting of superimposing multiple pairs of conformations are usually compiled into a distance matrix expressing the similarity between all the conformations in a data set. In a limited number of cases the distance matrix in itself is a useful tool. A recurrent example is a (dis)similarity with a block diagonal form, indicating the number of energy transitions possible to the molecule. However, normally this matrices do not adequately reflect the underlying conformational space [Becker 1997(b)].

Up until now we have not given much attention to the calculation of the RMSD values for the simulated trajectories because this would imply the choice of a dendrimer reference structure on which to superimpose all the others. As there are not any guidelines for the choice of such structure, this would ultimately consist on an arbitrary choice. Thus we have tried to avoid this problem as far as possible.

Moreover, calculating the RMSD for dendrimer structures using the presently available computational tools is not an obvious task. In fact, determining the RMSD

values for two superimposed dendrimer structures is hindered by a specific problem, which has its origin in the radial and quasi-symmetrical distribution of the dendrimers constituents.

RMSD and other distance-based similarity measures were initially developed to study proteins, which are basically linear compositions of consecutive amino acid residues. When applying these concepts to dendrimers we must keep in mind that these are branched macromolecules, and that, although we have defined a convention for the numbering of residues based on linear peptide chains, some chains will encompass more residues than others (see Figure 2.6). This is the problem associated with defining and cataloging branched symmetrical (or quasi-symmetrical) chemical structures. And it is impossible to ignore it because almost all tools/programs that resort to physics or mathematics need a way to sequentially differentiate the input.

In the case of dendrimers the ideal situation would involve a chain nomenclature able do discriminate between residues assigned to the same relative position while considering them as numerical equivalents.

To overcome this problem we devised an automated procedure to construct all meaningful dendrimer atomic-position combinations, by iteratively permuting residue sections bounded to the same branching units. This corresponds to exchange the positions of equivalent residues.

Accordingly we will have seven “hot spots” were permutations will iteratively occur. These correspond to the branching residues.Their atomic positions will remain the same while the atomic positions of the other residues will change (with exception of the core residues that do not have a corresponding set of residues with which to switch). This is illustrated in Figure 19 for the case of a simple two-dimensional tree.

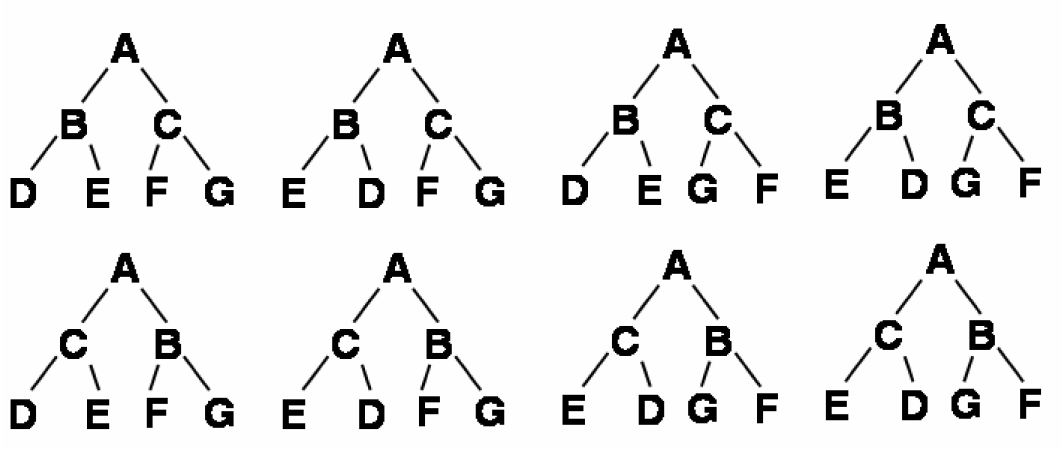


Figure 3.19 – The 2^3 = 8 permutations of a two-dimensional perfectly symmetric tree with 3 branching points.

Therefore, for any dendrimer there will exist 2^B permutations, where B corresponds to the number of branching points (number of bifurcations). In our case for every conformation obtained there will exist 2^7=128 equivalent atomic configurations, obtained by permutation of the spacer and peripheral residues among them.

Only by superimposition all 128 permuted conformations obtained for a single conformation with the reference conformation, and calculating all the 128 possible RMSD values, can we choose the minimum value, that will correspond (by definition) to the "true" RMSD value.

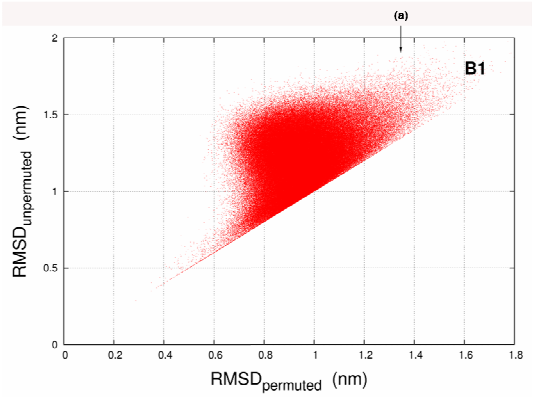
Thus, the use of permutations ensures that all appropriate residue-number combinations are accounted for and the problem associated with a meaningful RMSD calculation is surpassed.

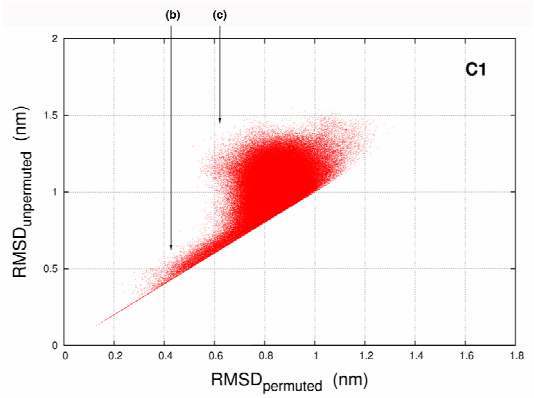
Due to the high computational cost associated with the calculation of RMSD matrices, we have performed it using only the conformations present at 1 ns intervals of the equilibrated trajectories for each replicate. This accounts for a total of 860 (B1),915 (B1H), 915 (B1HH), 890 (B1HHH) and 845 (C1) reference structures, each with its own 128 permuted conformations.

The final matrix containing the lowest value found between every pair of permuted conformations will have Y^2 entries (or elements), where Y corresponds to the total

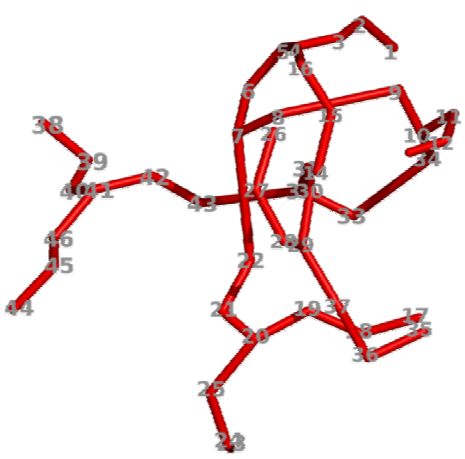
number of input structures. In practice this is equivalent to determine the 128 individual matrices and choosing the lowest value of each element to build a final, *a posteriori*, lowest value matrix.

The program used to calculate the RMSD values was *g_rms* and we have used all atoms in the conformations. To construct the PDB input files of the permuted conformations and the final matrices, we have developed specific programs. The program responsible for generating the permutations was named *permute_dendrimer* and the concepts behind its development were based on *tree representations*.[36]





(a)

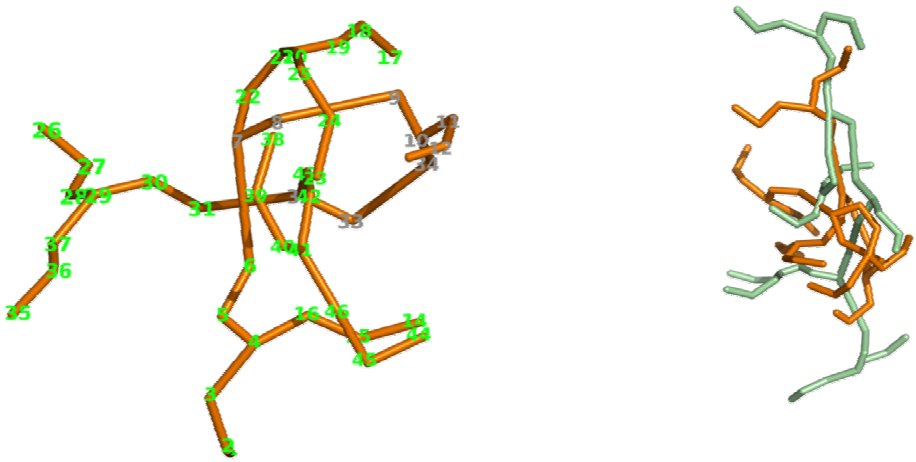


RMSD$_{unpermuted}$ = 1.998

Continues on next page

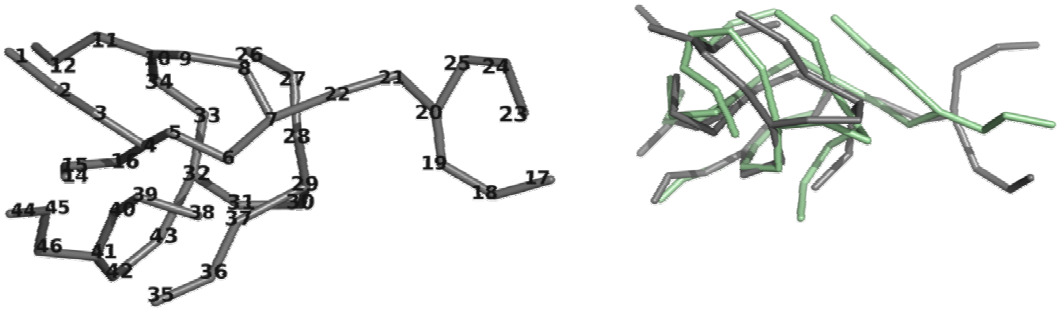
[36] The *tree* concept employed were refers to the definition used in graph theory. A tree is a graph in which any two vertices are connected by exactly one simple path.

(a)

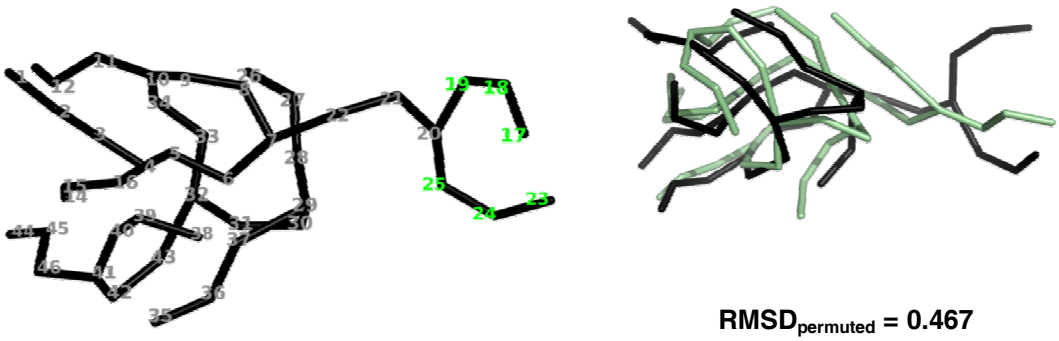


$RMSD_{permuted}$ = 1.386

(b)



$RMSD_{unpermuted}$ = 0.530



$RMSD_{permuted}$ = 0.467

Continues on next page

(c)

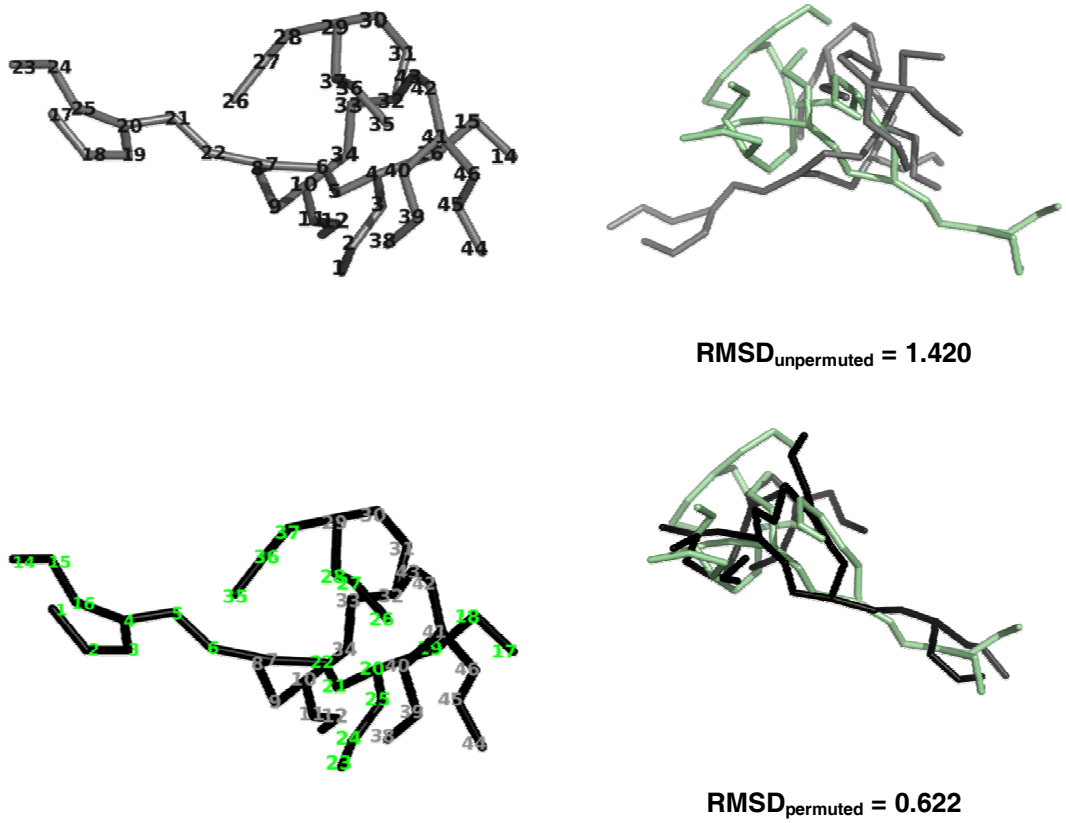


Figure 3.20 – Permuted versus unpermuted RMSD values.

Relation between the RMSD values obtained "normally" and using a permutation approach over all the residues segments between branching units. The fits were performed over all pairs of conformations within a set of structures chosen from the trajectories of B1 and C1. See the text for details.
(a),(b) and (c) exemplify the improvements promoted by using a permutation methodology. The images on the left correspond to the unpermuted and the "homologue" permutation with the lowest RMSD value after superimposition to the same structure; the superimpositions and related RMSD values are presented on the right. **(a)** B1 structure; **(b)** and **(c)** C1 structures.

In Figure 3.20 we plot the RMSD values obtained for each conformation pair, using both the unpermuted and permuted RMSD determination approaches.

As can be seem, the values differ, and hence the permutation over some residues normally originates structures that are chemical equivalents of the "original" one but present a greater similarity with the reference structures. If the permutations had not

improved the results, all the points in the plots would be overlapping at the $RMSD_{unpermuted} = RMSD_{permuted}$ (f(x)=x) diagonal line.

Furthermore, when comparing the plots obtained for B1 and C1 we verify that the magnitude of the RMSD changes is different. The RMSD values variation induced by the use of permutations is more frequent in B1. One can speculate that this probably happens due to the more symmetrical nature of Dap by comparison to Lyr (see Figure 2.5).

We have also selected three interesting data points (a, b, and c) from the RMSD comparison plots to illustrate the success of permutations, and different interesting situations.

In Figure 3.20(a) we observe a situation were a very high $RMSD_{unpermuted}$ value greatly improves, and although it still corresponds to one of the higher $RMSD_{permuted}$ values, the magnitude of the value is much smaller. The figure depicts the structures conformations with all residues (except the terminal amine) numbered.

In the top, Figures 3.20 (a)(b)(c) show the unpermuted conformations, with the residues numbers highlighted and the corresponding RMSD superimposition to the reference structure, thus obtaining the value for $RMSD_{unpermuted}$; below the figures present the structures with the permuted residues number highlighted in green and the unpermuted ones in the same color as in the top illustration. The superimposition to the reference structure and the correspondent RMSD values are also presented.

In Figure 3.20 (b) we represent conformations corresponding to a RMSD value that presents a small change with the use of permutations. Figure 3.20(c) represents a permutation induced RMSD-variation that alters the RMSD value to less than half the unpermuted one.

With the calculated RMSD values we have built the corresponding RMSD matrix that account for the (dis)similarity when superimposing pairs of all structures. Using this matrix we have determined the RMSD *central structure* for each dendrimer. These structures are presented next.

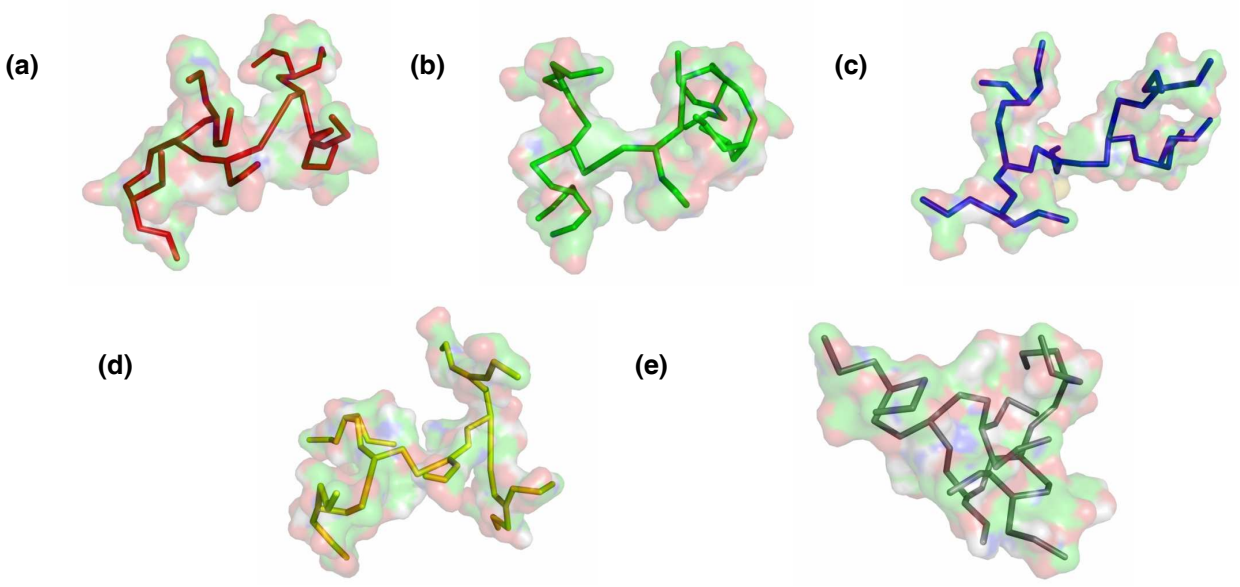


Figure 3.21 – Peptide dendrimers central structures.

Structures that minimize the RMSD between all pairs of structures within a particular conformational ensemble. See the text for details. **(a)** B1; **(b)** B1H; **(c)** B1HH; **(d)** B1HHH; **(e)** C1.

The *central structure* [Campos 2009] corresponds to the structure (*A*) chosen among all the available *N* structures, that minimizes the dispersion measure, D:

$$D_A^2 = \frac{1}{N-1}\sum_{B=1}^{n} RMSD_{AB}^2 , \qquad \text{Equation 3.7.}$$

where $RMSD_{AB}$ is the RMSD value between structures *A* and *B* considering all their atoms.

With the *central structure* we now possess a well defined, meaningful reference structure on which to superimpose all other structures, and thus obtain RMSD values that can effectively be used, not only as a dissimilarity measure between sets of conformations, but also as a reaction coordinate for the system. A consequence of using the *central structure* as the reference one, is that within the set of sampled conformations, there exits at least one structure that has a RMSD value equal to zero.

3.2.8. RMSD and R_g free energy landscapes

By now, we have already concluded that dendrimer folding is a complex problem, and that it will probably require several coordinates for a detailed description. This arises from the myriad of dendrimer intra-convertible conformations that we have established to exist. Thus, we have performed a more thorough analysis of the free energy landscapes along several folding coordinates [Brooks 2002; Guo 1997; Shea 2001]. We expect that by exploring these multidimensional free energy surfaces of peptide dendrimers we can shed some light on the statistics underlying their folding.

To this end, we have computed the free energy as a function of two reaction coordinates: the radius of gyration (R_g) and the RMSD value obtained for a fitting using as reference the *central structure* ($RMSD_{cs}$). We have used these two coordinates no only because of the easiness to compute them, but also because with this approach we are calculating a conditional free energy as a function of an overall measure for compactness (that is tightly related to the shape of dendrimers), and a relative measure for the similarity among structures.

The resulting free energy hypersurfaces are used to investigate the existence of conformational classes by determining their energy minima.

The potential of mean force[37], or conditional free energy projected onto the reaction coordinates (R_g and $RMSD_{cs}$), can be computed as:

$$E(R_g, RMSD_{cs}) = -RT \ln \frac{P(R_g, RMSD_{cs})}{P\max}, \qquad \text{Equation 3.8.}$$

where $P(R_g,RMSD_{cs})$, P_{max}, T, and R are respectively, the probability density function, its maximum, the absolute temperature, and the ideal gas constant[38] [Campos 2009, Shea 2001].

[37] An elegant way of gaining insights into the folding of molecules is through the generation and study of free energy as a function of reaction coordinates describing the molecule folding progress. The Potential of Mean Force (PMF) is an effective configuration-dependent free energy potential that makes no explicit reference to the solvent degrees of freedom [Becker 2001].

The probability density function in the representation space was estimated from the ensemble of conformations using a Gaussian kernel estimator; an approach identical to the one performed elsewhere [Campos 2009].

We computed the conditional free energy using 4251 (B1), 4526 (B1H and B1HH), 4401 (B1HHH) and 4176 (C1) structures extracted from the concatenated trajectories at intervals of 0.2 ns. The list of programs used in these calculations include: *getdensity*[39], *g_rms*, *permute_dendrimer*, *g_gyrate*, and a number of in-house scripts.

Over the next figures, we examine the distribution of distinct conformations, $P(R_g,\mathrm{RMSD}_{cs})$, sampled during the folding process. The progress of folding is measured based on the used reaction coordinates. The free energy hypersurfaces for each dendrimer are displayed as contour maps, which indicate how the free energy varies with changes in the compactness (R_g) and the similarity to the central structure (RMSD_{cs}). In the figures we have also represented a scatter plot illustrating the correlation among the used folding-progress coordinates.

[38] The ideal gas constant, R, is equivalent to the Boltzmann constant, k_B, but it is expressed in energy units ($R = N_A k_B$, where N_A is the Avogadro number).

[39] *Getdensity* is a Sara Campos and António Baptista development for computing probability density distributions and free energy values [Campos 2009].

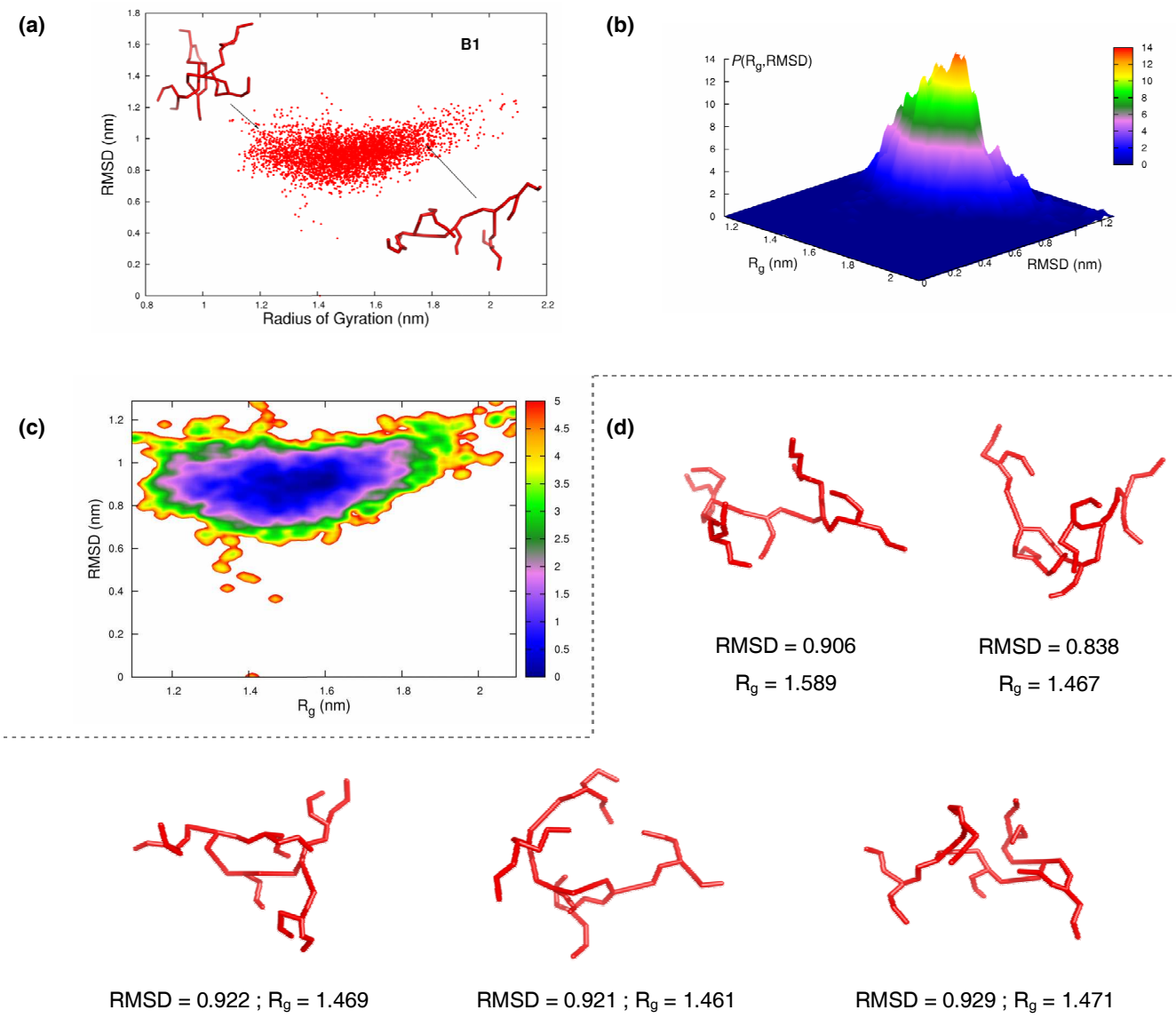


Figure 3.22 – B1 energy surface using R_g and $RMSD_{cs}$ as folding coordinates.

(a) Scatter-plot relating R_g and $RMSD_{cs}$ values of each conformation. The structure on the top left corner of the scatter plot has R_g = 1.236 nm; $RMSD_{cs}$ = 1.001 nm. The structure on the right bottom corner of the scatter plot has R_g = 1.805 nm and $RMSD_{cs}$ = 1.001 nm. **(b)** 3D surface exhibiting the probability density values. **(c)** Conditional free energy landscapes (projected on 2D) of B1 using as reaction coordinates R_g and $RMSD_{cs}$. **(d)** B1 conformations located at, or near, the energy minimum. The RMSD values presented in the figure are $RMSD_{cs}$, meaning that they were computed by superimposition to the central structure. The lowest energy conformation is the first one on the left (R_g = 1.589 nm; $RMSD_{cs}$ = 0.906 nm). The units for both the R_g and $RMSD_{cs}$ values are nanometers. The scale presented for the energy contour is in RT units.

In the figure, we present both the contour for the free energy surface and the probability surface for the data. All the structures were represented centered in the core, with the first bifurcation horizontally oriented. This should make a comparison easier.

We start by analyzing the scatter plot illustrating the data correlation. The plot suggests that there is no obvious relationship between the RMSD and the R_g values. Still, and as illustrated by the adjacent structures, some points with similar values of $RMSD_{cs}$ represent very distinct conformations. As such, we can consider that the RMSD using as reference the *central structure* is not, by itself, a great reaction coordinate. However, this does not implies that a conformational space comprising this RMSD values and some other reaction coordinate (such as R_g) suffers from the same problem.

The three-dimensional probability surface, although with some general roughness, illustrates that, prior to "find" its lowest energy value, the dendrimer samples only a very small set of distinct conformations, all separated by small energy barriers. This idea is further supported by the free energy plot. As evident from the contour figure, the B1 conditional free energy has a general funnel-like form, implying that the folding process is essentially downhill in free energy. Furthermore, we observe only one energy basin, located at $R_g \approx 1.59$ nm and $RMSD_{cs} \approx 0.90$ nm. The *basin* encompassed, not only, the lowest energy structure (global minimum), but also the more "expanded" energy regions.

We note that the energetic transitions among the different conformations are characterized by very small energy values. In fact, to visualize the representative parts of the surface we had to impose a maximum *cutoff* of E = 5 RT for the free energy values. We consider the *basins* within this *cutoff* as conformational *classes*. Excluding high free energy conformations through the imposition of a *cutoff* value promotes the homogeneity of conformational classes, and is only possible because those conformations are more likely to present structural deviation from the conformation in the respective minima. Moreover, due to the method used, the high-energy regions of the hypersurface are less reliable than the low-energy ones, because they are more exposed to sampling problems [Campos 2009].

3. Results and Discussion

Looking at some of the structures collected from the only existing *basin* we verify that they share some features; all present a mildly loose structure, and none of them presents interlacing branches (*bowl-like shapes*). A clear distinction between the structures is not obvious.

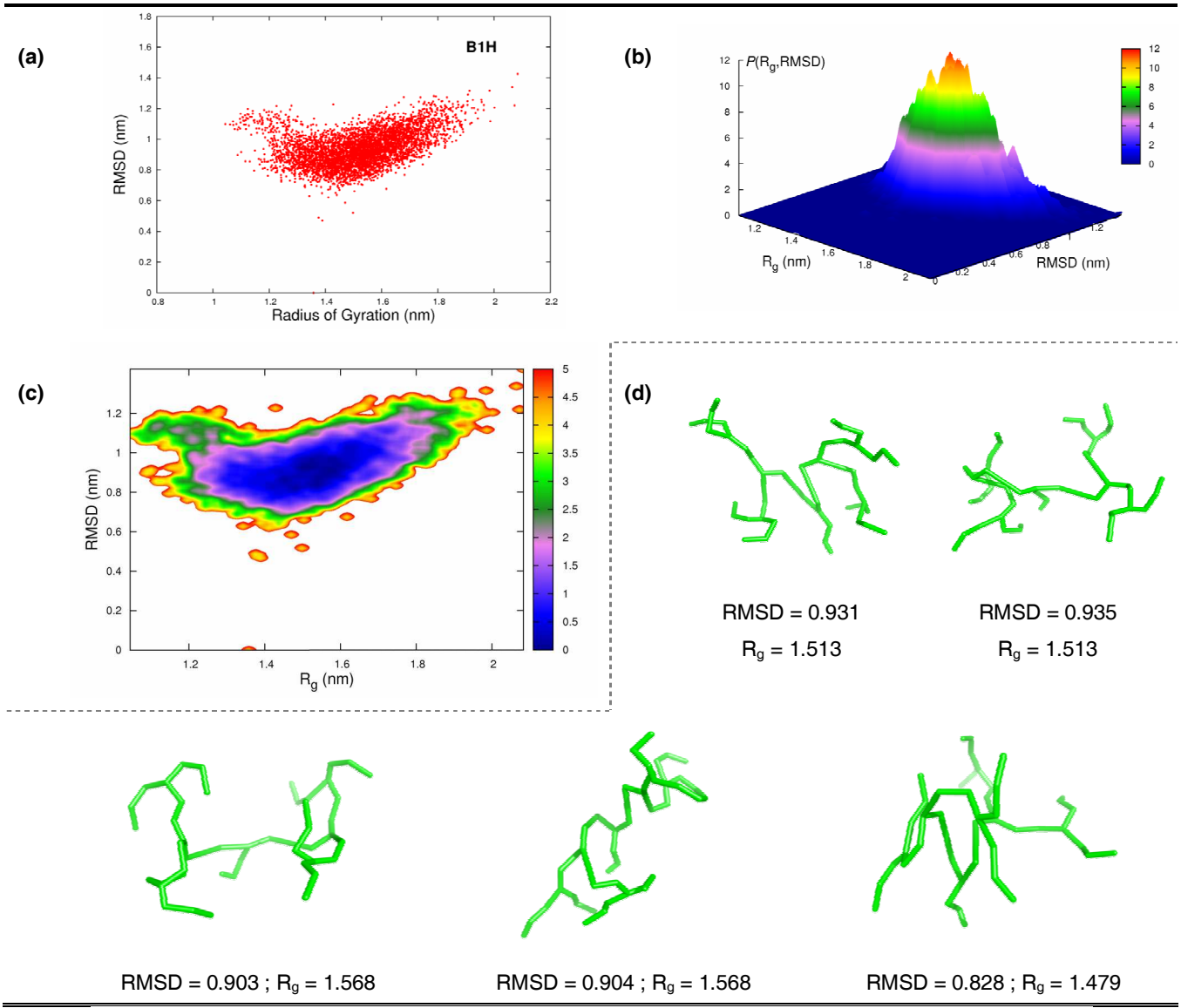


Figure 3.23 – B1H energy surface using R_g and $RMSD_{cs}$ as folding coordinates.

(a) Scatter-plot relating R_g and $RMSD_{cs}$ values of each conformation. **(b)** 3D surface exhibiting the probability density values. **(c)** Conditional free energy landscapes (projected on 2D) of B1H using as reaction coordinates R_g and $RMSD_{cs}$. **(d)** B1H conformations located at, or near, the energy minimum. The RMSD values presented in the figure are $RMSD_{cs}$, meaning that they were computed by superimposition to the central structure. The lowest energy conformation is the first one on the left (R_g = 1.513 nm; $RMSD_{cs}$ = 0.931 nm). The units for both the R_g and $RMSD_{cs}$ values are nanometers. The scale presented for the energy contour is in RT units.

The suggestion following from the observation of Figures 3.23 and 3.24, is that, some of the conclusions redrawn from the analysis of B1 are generally applicable to all the members of the B1-family dendrimers. Accordingly, the relationship between the R_g and $RMSD_{cs}$ values, explicit in the scatter plots, is similar for all B1-family dendrimers. Furthermore, the three-dimensional probability surfaces always present some roughness, but are generally indicative of downhill folding; all present a single energy minimum; the images of structures that accompany the contour plots suggest the mixing of similar conformations.

Furthermore, analogously to B1, in all of the other B1-family dendrimers, the structures that correspond to the free energy minima have minimal (if any) inter-branch contacts (regardless of the generation considered), and seem to favor a bowl-like geometry.

B1H specifically, presents one global minimum, located at $R_g \approx 1.51$ nm and $RMSD_{cs} \approx 0.93$ nm.

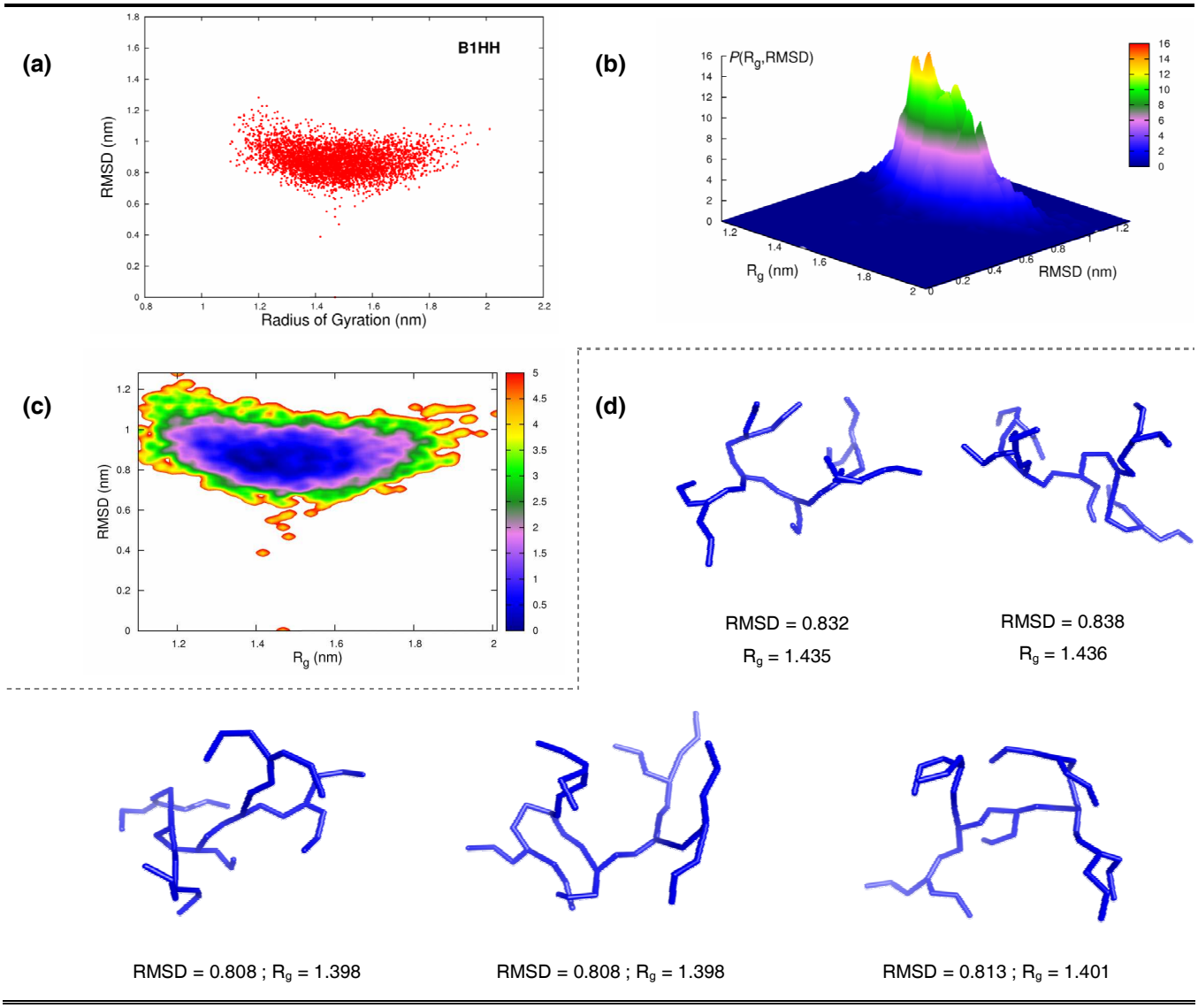


Figure 3.24 – B1HH energy surface using R_g and $RMSD_{cs}$ as folding coordinates.

(a) Scatter-plot relating R_g and $RMSD_{cs}$ values of each conformation. **(b)** 3D surface exhibiting the probability density values. **(c)** Conditional free energy landscapes (projected on 2D) of B1HH using as reaction coordinates R_g and $RMSD_{cs}$. **(d)** B1HH conformations located at, or near, the energy minimum. The RMSD values presented in the figure are $RMSD_{cs}$, meaning that they were computed by superimposition to the central structure. The lowest energy conformation is the first one on the left (R_g = 1.435 nm; $RMSD_{cs}$ = 0.835 nm). The units for both the R_g and $RMSD_{cs}$ values are nanometers. The scale presented for the energy contour is in RT units.

The free energy surface for B1HH reflects a global minimum located at $R_g \approx 1.43$ nm and $RMSD_{cs} \approx 0.83$ nm.

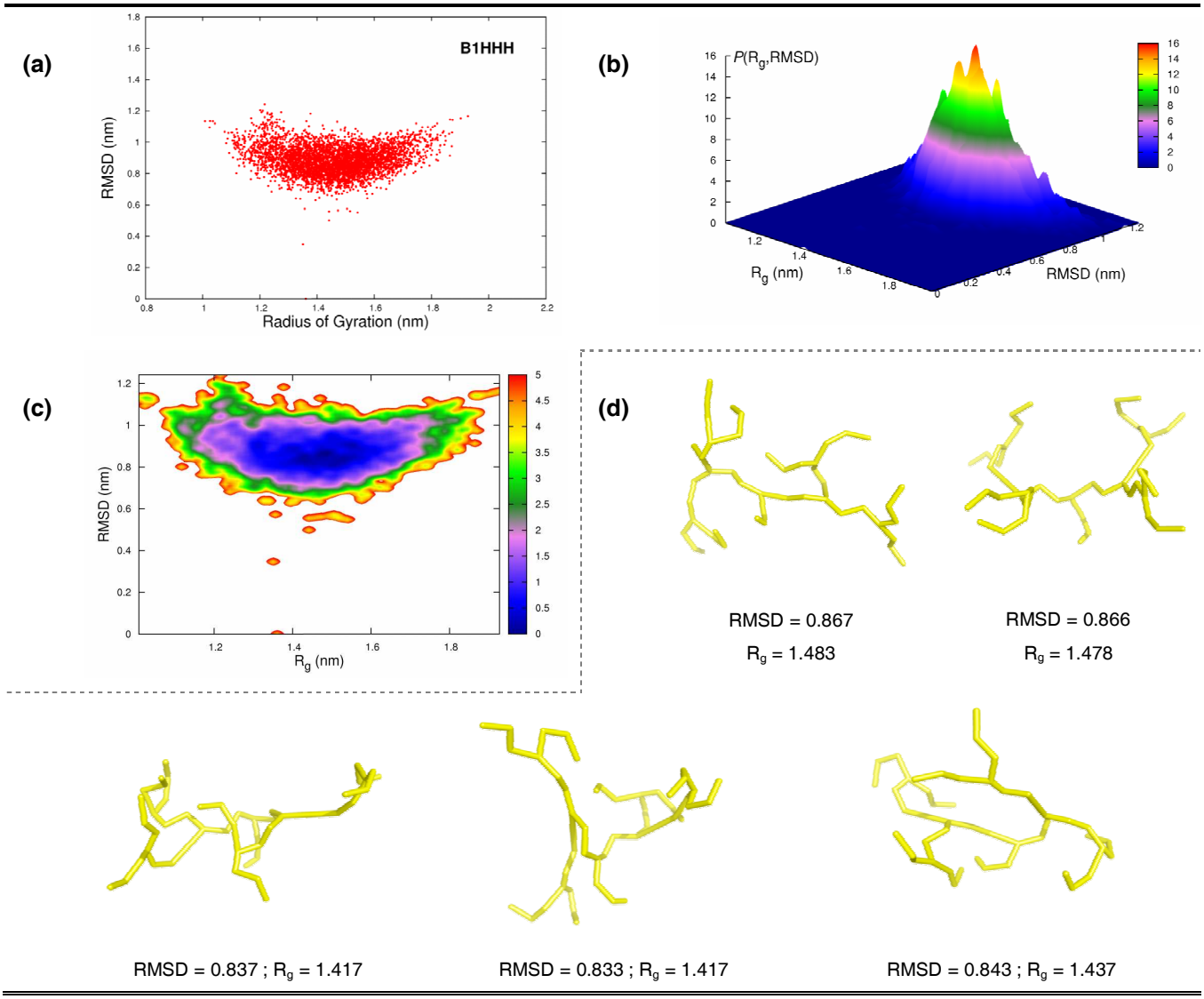


Figure 3.25 – B1HHH energy surface using R_g and $RMSD_{cs}$ as folding coordinates.

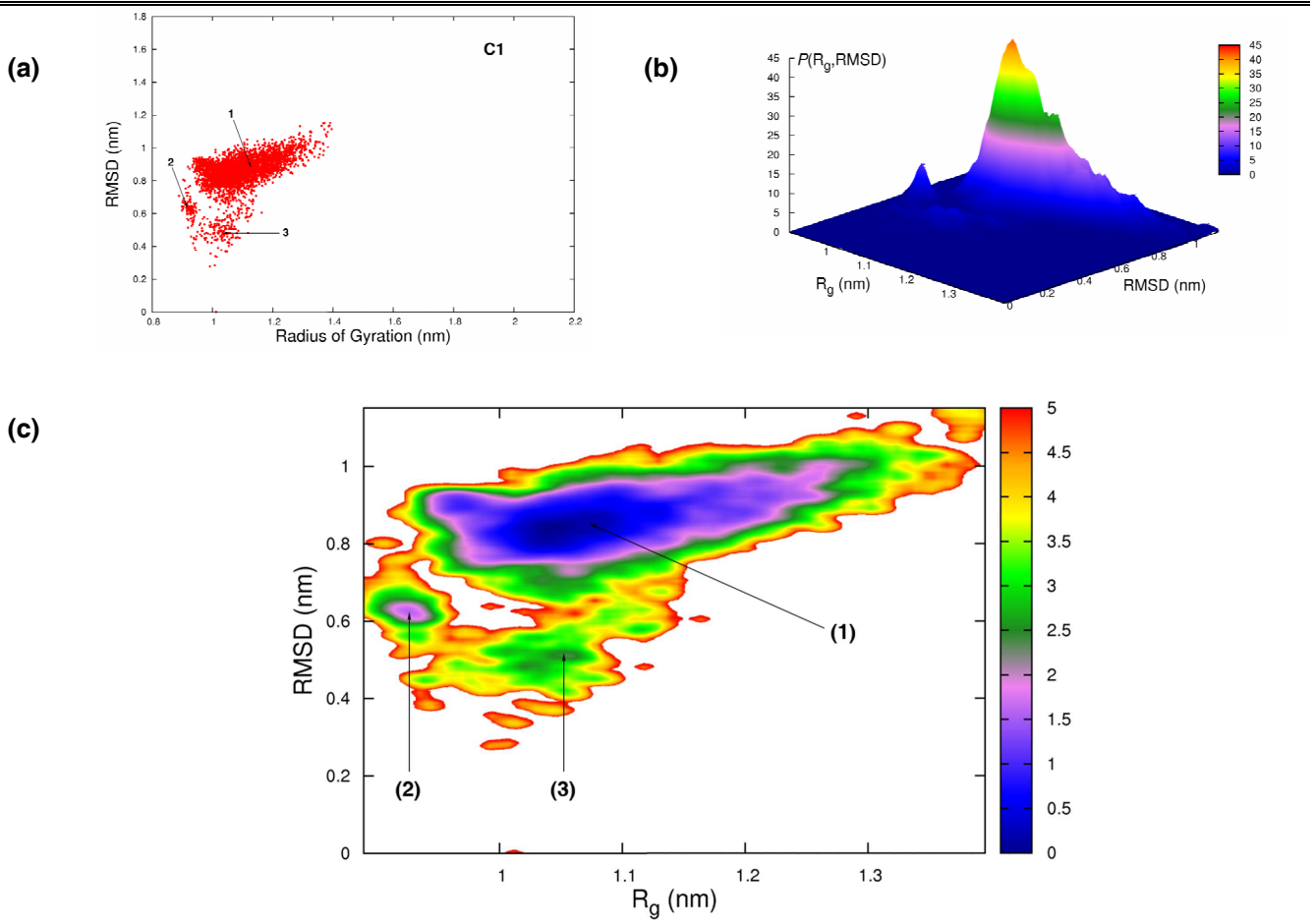
(a) Scatter-plot relating R_g and $RMSD_{cs}$ values of each conformation. **(b)** 3D surface exhibiting the probability density values. **(c)** Conditional free energy landscapes (projected on 2D) of B1HHH using as reaction coordinates R_g and $RMSD_{cs}$. **(d)** B1HHH conformations located at, or near, the energy minimum. The RMSD values presented in the figure are $RMSD_{cs}$, meaning that they were computed by superimposition to the central structure. The lowest energy conformation is the first one on the left (R_g = 1.483 nm; $RMSD_{cs}$ = 0.867 nm). The units for both the R_g and $RMSD_{cs}$ values are nanometers. The scale presented for the energy contour is in RT units.

The contour plot for B1HHH also suggests downhill folding. The energy minimum is placed near $R_g \approx 1.48$ nm and $RMSD_{cs} \approx 0.86$ nm.

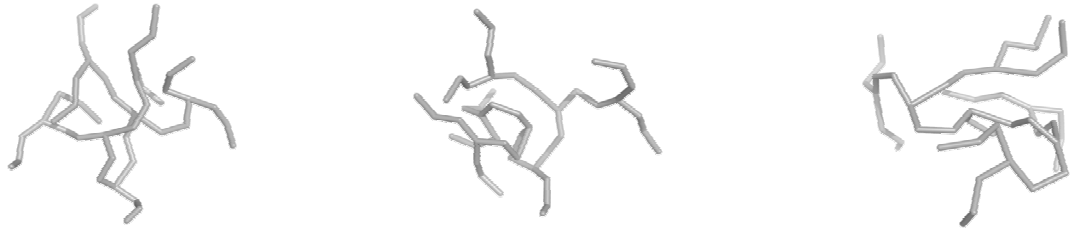
3. Results and Discussion

After looking at all the B1-family energy contours, and considering their overall features, we detect similarities between the shapes and the energy regions of all contour plots. These findings support the idea that, in terms of energetic barriers, all B1-family dendrimers behave similarly, regardless of the amino acid differences that characterize each of them. It seems that a small number of amino acid changes in the composition of a peptide dendrimer do not account for major transformations in the energy landscapes defined using R_g and $RMSD_{cs}$ as reaction coordinates.

Next we present the same approach using C1.



(1)



RMSD = 0.843 ; R_g = 1.044 RMSD = 0.848 ; R_g = 1.042 RMSD = 0.844 ; R_g = 1.045

Continues on next page

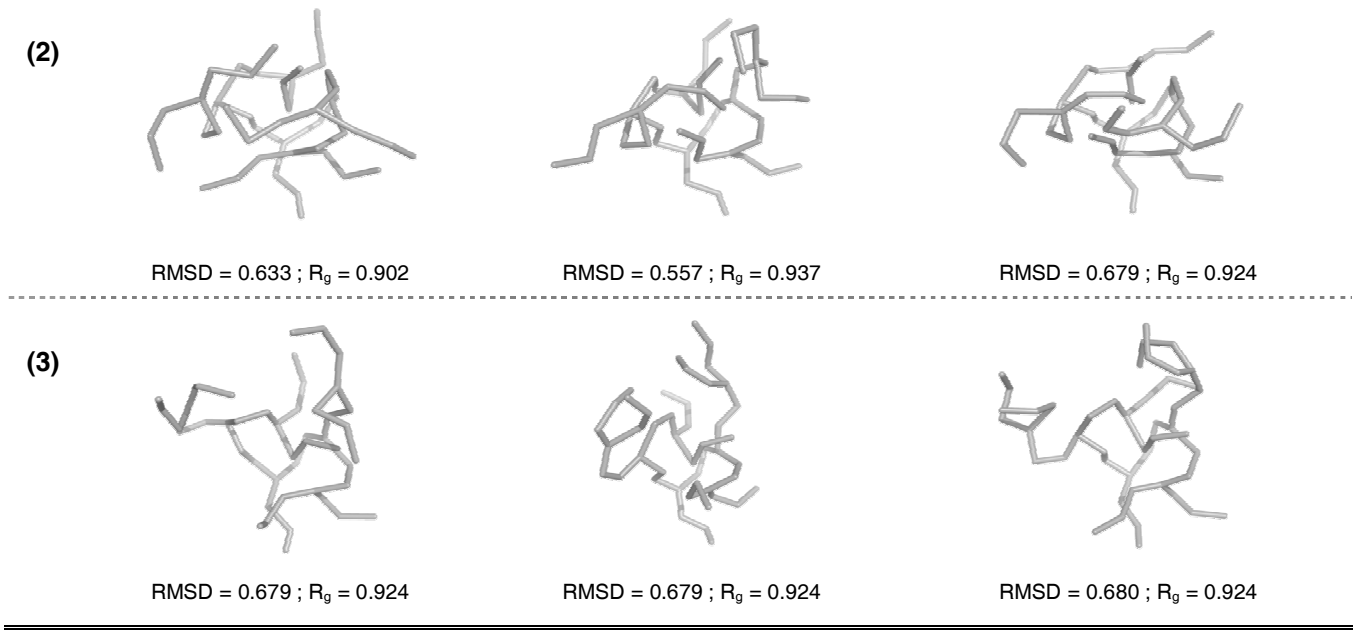


Figure 3.26 – C1 energy surface using R_g and $RMSD_{cs}$ as folding coordinates.

(a) Scatter-plot relating R_g and $RMSD_{cs}$ values of each conformation. **(b)** 3D surface exhibiting the probability density values. **(c)** Conditional free energy landscapes (projected on 2D) of C1 using as reaction coordinates R_g and $RMSD_{cs}$. **(1)(2)(3)** C1 conformations located at, or near, the energy minima. The RMSD values presented in the figure are $RMSD_{cs}$, meaning that they were computed by superimposition to the central structure. The lowest energy conformation is the first one on the left (R_g = 1.044 nm; $RMSD_{cs}$ = 0.843 nm). The units for both the R_g and $RMSD_{cs}$ values are nanometers. The scale presented for the energy contour is in RT units.

Some significant differences are visible in the B1-family and C1 scatter plots. The scatter plot for C1, contrary to the point mantle in B1-family dendrimers, has some visible point agglomeration. We observe three distinct groups of points (we have numbered them as 1, 2 and 3, see the corresponding figure). The first, **1**, congregates most of the data, as such, it should be a very heterogeneous cluster. The second group, **2**, and the third, **(3)**, gather a reduced number of structures.

The question remaining is whether different structures presenting distinct features are gathered in these groups, thus defining the designed conformational space as a valid one for conformational analysis of C1.

By analyzing the three-dimensional probability surface for C1 it is clear that, although it presents some roughness, it mostly shows an unequivocal uphill tendency (in the probability, that is roughly the inverse of the free energy), where, despite the existence of two peaks, one of them clearly dominates in number of conformations. However, the existence of two clearly separated probability lumps might suggest an effective cluster separation. To check this hypothesis we must analyze the free energy surface.

As indicated by the scatter-plot, we are able to observe three distinct energy wells. However, one of them is characterized by much higher number of high-energy conformations (pinpointed as **3** in the contour map) than the others, **1** and **2**. In fact, the conformations accounted for by **3** constitute region characterized by energy values varying between 5 RT < E(R_g,RMSD$_{cs}$) < 2.5 RT. Observing some of the structures present in this "cluster" we conclude that they share a clear structural pattern. These structures present sphere-like shapes with a packed geometry and a similar branching distribution. Furthermore, the free energy map suggests that, transitions between this "cluster" and the minimum energy one, would have to overcome very high energy barriers, in the order of several multiples of RT.

The energy basin pinpointed as **2** in the contour map account for conformations with free energies of approximately 1.75 RT; displaying very compact structures that seem to favour contacts between the branches, an thus forming a condensed folded nucleus.

In terms of low energy we observe a single energy minimum near $R_g \approx 1.04$ nm and RMSD$_{cs} \approx 0.84$ nm where the existent structures share some features but are still inhomogeneous. Nonetheless, their overall shape is very similar, presenting compact/condensed structures. A curious and noteworthy occurrence concerns the fact that the structures presented in the clusters (**2** and **3**), are more compact (see the R_g values). This implies that the preferential configurations of C1 are not the most compact ones.

The low-dimensional free energy landscapes obtained (for all dendrimers) are mainly intended as graphical guides and support because we will have to resort a more detailed analysis using a different angle to confirm these results.

3.2.9. Peptide dendrimers size and shape

Until now we have discussed the spatial configuration of dendrimers as a function of some of their properties and their distributions according to some reaction coordinates, this has proven to be inconclusive.

Having performed a extensive analysis of some folding coordinates we now turn our attention to a different and more unusual approach, the shape analysis.

In this section, we will present a shape analysis for simulated peptide dendrimers. This analysis is mainly associated with the quantification of asymmetries and deviation of the internal molecular axes in regard to some defined symmetries. Thereby, introducing some basic concepts and defining some "shape" measures is mandatory before proceeding.

Obviously the molecular structure has a marked effect upon the shape of a molecule. Šolc and Stockmayer [Solc 1973] were the first to explicitly discuss the shape of molecules (specifically polymers). In their work they recognize that R_g characterizes the distribution of atoms about the center of mass for a specified conformation of a molecule. As such, they introduced as shape measures the eigenvalues of the *radius of gyration tensor*[40].

These shape parameters characterize the instantaneous state of a particular molecule by more than just its size. For these purpose we use the three eigenvalues of the *radius of gyration tensor* (S) deduced from the principal orthogonal components (L_1, L_2 and L_3) of its square radius (R_g^2), by decomposing R_g^2 along the principal axes of inertia of the molecules (**S**, is the diagonalized tensor). The principal components L_n^2 (n=1,2,3) determine the size and the shape of an ellipsoid which is equivalent, in regard to its inertial properties, to the macromolecule and they can be thus considered as the shape parameters of the molecule as well.

[40] The nomenclature regarding shape descriptors is not consensual. In here we will employ the terminology defined by Šolc [Solc 1973] and thoroughly explained by Mattice [Mattice 1994].

Herein our interest will focus essentially on some internal relations between these components, and their relation to R_g. We will determine some quantities normally associated with shape analysis (*shape descriptors*) and evaluate their relation with the conformational variability of peptide dendrimers.

The measures computed are the *asphericity* (*b*) and *acylindricity* (*c*) [Theodoru 1985; Steinhauser 2005; Mattice 1994]. These quantities are defined using the elements of **S** diagonal as

$$b = \mathrm{L}_3^2 - \frac{1}{2}\left(\mathrm{L}_2^2 + \mathrm{L}_1^2\right), \qquad \text{Equation 3.8.}$$

and

$$c = \mathrm{L}_2^2 + \mathrm{L}_1^2 \,. \qquad \text{Equation 3.9.}$$

Some considerations and details on shape analysis, and the parameters can be found elsewhere [Mattice 1977; Mattice 1994; Solc 1973; Zifferer 1999].

As the name states, asphericity and acylindricity can be roughly conceived as measure for the structural difference to a sphere and to a well defined cylinder respectively. Thus, shape parameters might be useful conformational descriptors (alone or in conjunction with some of the previously discussed reaction coordinates) for discussing the existence of conformational *classes*. The approach we will employ here is similar to others performed elsewhere [Steinhauser 2005; Zifferer 1999]. However, and to the best of our knowledge, this approach has never been performed for peptide dendrimers.[41]

Most shape analysis base their results on coefficients between the eigenvalues of **S**. Here, we will avoid this methodology and employ a more intuitive one (in our opinion). We start by determining the values for L_1, L_2 and L_3. The set of values calculated for each conformation are then used to define a three-dimensional space

[41] We note that although this approach has never been performed for peptide dendrimers, the study and evaluation of dendrimers size an shape, has already been achieved for several dendritic systems, especially PAMAM type [Han 2005; Maiti 2005; Lee 2006].

where each eigenvalue is one of the system coordinates (see Figure 3.27(a)). For the sake of simplicity, we have named this 3D space as *gyration space* (G_{space} for short).[42]

With this representation we expect to create a clear picture of the symmetry axes (and shape preferences) within the different conformations. Moreover, using this space we can represent on it the different peptide dendrimers simultaneously, thus obtaining a comparative notion for the shape of each of them.

The calculation of the eigenvalues and associated *b* and *c* values was carried out for an ensemble of conformations constituted by all the structures collected during MD in equilibrium. Accordingly, the ensembles are constituted by 85010, 90510, 90510, 88010 and 83510 conformations for B1, B1H, B1HH, B1HHH and C1 respectively.

To compute the eigenvalues of **S** we have used *g_polystat*, a GROMACS tool [Berendsen 1995].

In Figure 3.27(a) we represent the values of L_1, L_2 and L_3 for B1 in the G_{space}. Furthermore, the figure displays the G_{space} values of B1 colored according to the asphericity and acylindricity of the corresponding conformations. Additionally, some structural examples are also presented.

[42] The Gyration Space is a "limited" space, not in the mathematical sense, but in the practical one. This statement is based on the fact that by definition the eigenvalues of **S** are always greater than or equal to zero. Therefore, the values of L1, L2 and L3 associated with a certain molecular conformation will always occupy a region comprehended within the first quadrant of a 3D space. Moreover, as we have defined $L_1 \leq L_2 \leq L_3$ there will exist a conceptual plane defined by L1=L2=L3 that corresponds to a diagonal plane intersecting the origin of the reference frame. More important, all the conformations will occupy a region of the space between that plane and the plane defined by the L_1,L_3 axes. All the conformations displaying values contained within the L_1=L_2=L_3 plane will represent a "perfect symmetry", where the existing shape is not of the corresponding ellipsoid, but of the corresponding sphere.

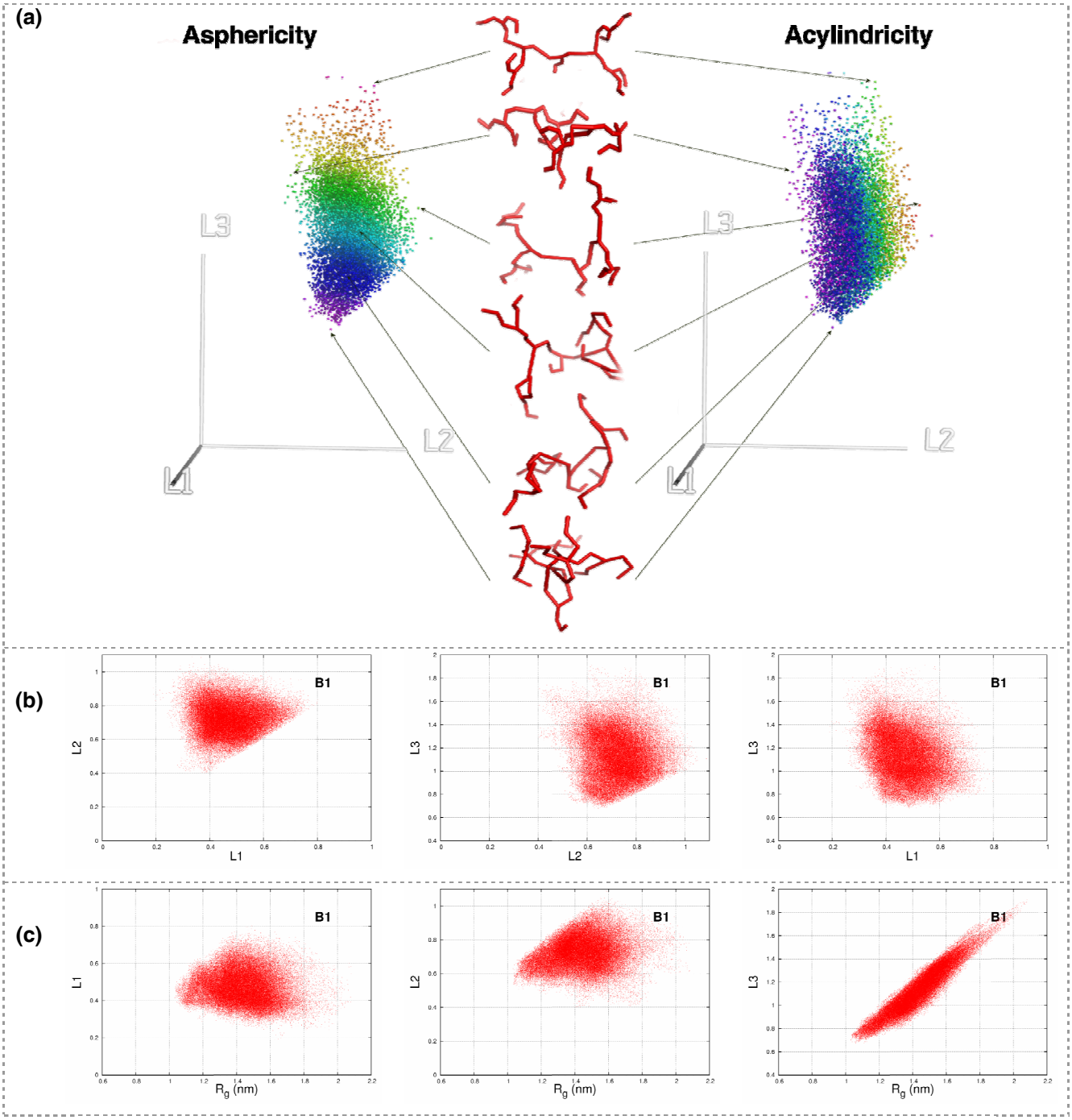


Figure 3.27 – Asphericity, acylindricity and gyration space representation for B1.

(a) Two identical views of the conformations distribution in gyration space, where the one to the left is colored according to the value of asphericity of each conformation; and the one to the right by the acylindricity values. The color scale represents the higher values as red and the lowest as violet. In the center of the figure are represented some structure examples. Due to the computational effort involved in constructing the 3D plot we have only represented a tenth of the total amount of points for which we have determined L_1, L_2 and L_3. (b) Relations among the different principal components of the radius of gyration tensor. (c) Relation between the radius of gyration and its diagonalized tensor components. For further details see the text.

The *b* and *c* 3D-G_{space} plots for B1H, B1HH and B1HHH are similar, in terms of occupied regions and of point distribution patterns, to the B1 3D-G_{space} plot. Therefore, the aforementioned plots are available for consulting in Appendix G. Scatter-plots on the relation between R_g and each G_{space} coordinate are also presented in that appendix.

As can be verified from the Figure 3.27, especially from the 2D-plots depicting the values of the possible eigenvalues combinations Figure 3.27(b), there are no conformations of B1 exhibiting "perfect" symmetry, i.e., with all the principal components of **S** sharing the same value (L_1=L_2=L_3). Despite that, we are able to identify conformations with pronounced symmetry, where two of the principal components (L_1 and L_2) share the same value. However, we have also confirmed the existence of conformations with very distinct values for the eigenvalues of **S**.

Since L_3 is the highest eigenvalue it is expected (as $L_1^2+L_2^2+L_3^2 = R_g^2$) to share a pronounced relation with R_g, being the eigenvalue that more adequately reflects, by itself, the dispersion of the data points. Within this rationale, L_1 is the eigenvalue where the relation to R_g is most tenuous. A quantitative corroboration of this expectations is presented in Figure 3.27(c).

Another implication of the relation between R_g and the principal moments of **S** is that, conformations closer to the origin of the referencd frame in *gyration space* must present smaller R_g values than the ones that are further apart. Accordingly, we can consider the existence of a R_g progressive gradient across the *gyration space*, where conformations nearer to the origin correspond to the more compact ones.

The distribution of conformations within G_{space} still holds some interesting information. For B1 (and all B1-family members) we find that the conformations occupy, in a consecutive and almost uniform manner, a significant portion of the space. Thus resulting in a tight mantle-like aggregate of points. Some extremity points are however, more "separated" from the main "mantle" (*outliers*). The fact is that no group (or cluster), apart from the dominant one, is visible in the representation of B1 conformations on G_{space}, implying that, relating to the shape/form of B1 conformations, it seems B1 is able to visit a multiplicity of different shape *categories* without privileging or circumventing any one in particular. This is specially emphasized by the images of structures in Figure 3.27(a), where we discern both

completely stretched, as well as shrunken conformations, connected by a succession of other differently shaped structures.

As for the asphericity and acylindricity of B1 conformations, we verify the existence of very different *c* and *b* values (see Figure 3.27(a)). The asphericity of B1 conformations becomes higher as the conformations begin to move away from origin of the reference frame. We also observe that the less aspherical conformations (and thus the more spherical ones) are mainly conformations with multiple interacting residues and inter-branch contacts, while the more aspherical ones are mainly the more loose conformations.

If asphericity seems to preferentially dispose its values "vertically", acylindricity shows the opposite behavior, distributing the data in a "horizontal" fashion. The structures presented in the figure help to elucidate the conformational characteristics associated with the acylindricity scale for B1. Contrary to *b*, the association between the disposition of points as a function of *c*, and the structural features of the associated conformations (evidenced by the images) is not very intuitive.

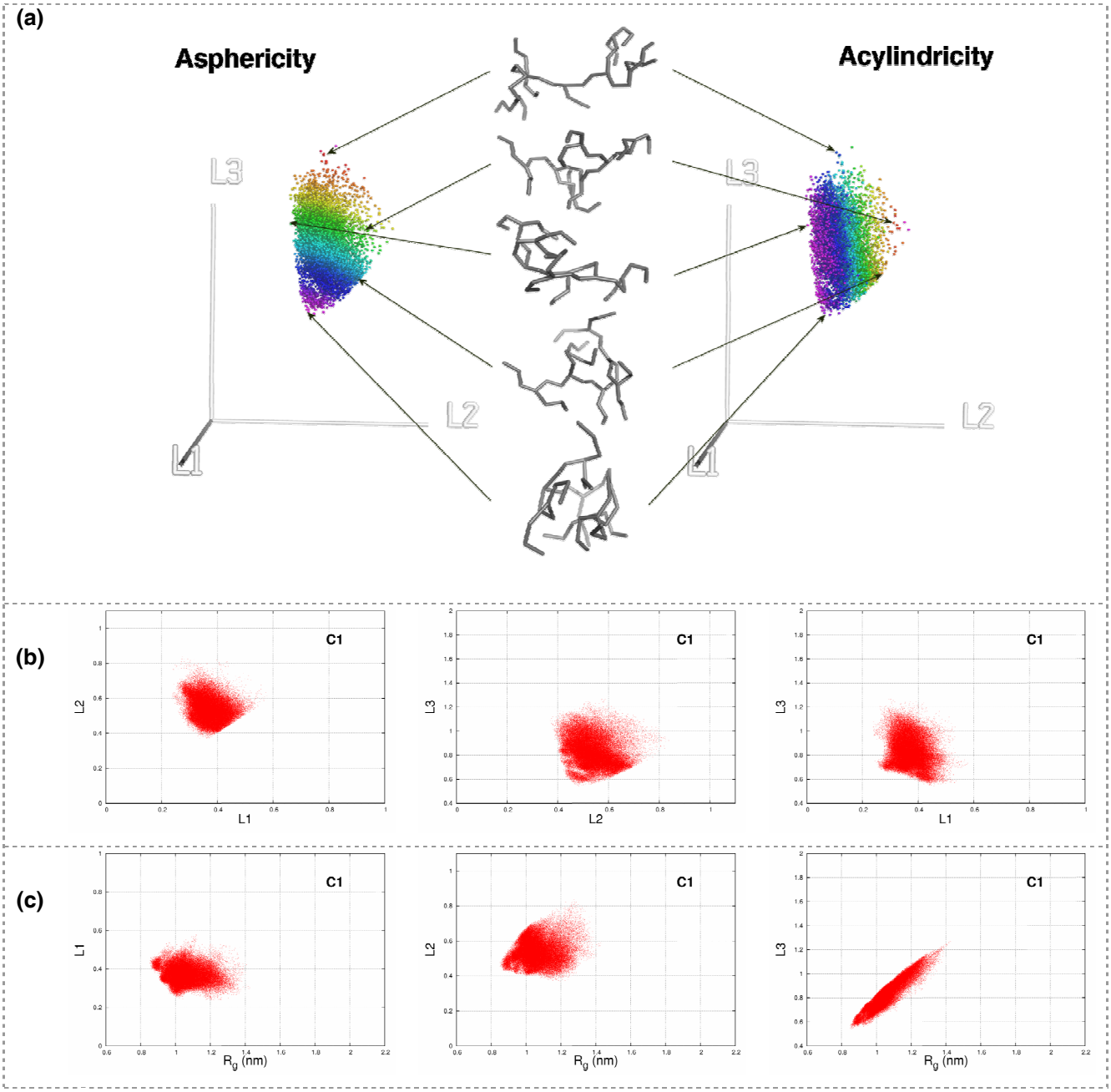


Figure 3.28 – Asphericity, acylindricity and gyration space representation for C1.

(a) Two identical views of the conformations distribution in gyration space, where the one to the left is colored according to the value of asphericity of each conformation; and the one to the right by the acylindricity values. The color scale represents the higher values as red and the lowest as violet. In the center of the figure are represented some structure examples. Due to the computational effort involved in constructing the 3D plot we have only represented a tenth of the total amount of points for which we have determined L_1, L_2 and L_3. **(b)** Relations among the different principal components of the radius of gyration tensor. **(c)** Relation between the radius of gyration and its diagonalized tensor components. For further details see the text.

The results obtained previously to the shape analysis of the peptide dendrimers suggested that C1 presented a spatial geometry of its own, with marked differences from all the B1-family dendrimers. This geometry is characterized by more compact structures with multiple interactions among residues in varied parts of the dendritic architecture.

Analyzing Figure 3.28, especially sections (a) and (b), we verify that the conformational trend of C1 has a correspondent trend in terms of the conformations shape. Given the fact that the plots in Figure 3.27 and Figure 3.28 are in the same scale, one can assume that C1 occupies a region of the gyration space featuring lower L_1,L_2 and L_3 values than any of the B1-family dendrimers (as was to be expected given the proven differences in terms of the R_g values magnitude).

Moreover, although C1 conformations demonstrated the same tendency as B1 to aggregate in a certain region of G_{space}, there is a higher number of *outliers*, and they are disposed in a more spaced manner; suggesting the possibility of existing shape *classes*. The existence and identification of such *classes* will be addressed briefly.

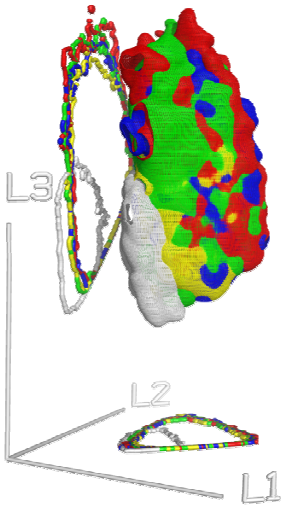
By now we just call the reader's attention to a noticeable set of low asphericity, low acylindricity, conformations near the origin of the referential, and that are visibly spaced from the bulk of points.

The images of dendritic structures presented in Figure 3.28(a) help to illustrate the *c* and *b* of some conformations placed at different regions of G_{space}. The conclusions that can be drawn from these figures are identical to the ones obtained for B1, except for the fact that in C1, the conformations connecting the highest and lowest asphericity conformations seem to be more compact and present more interacting residues than in B1.

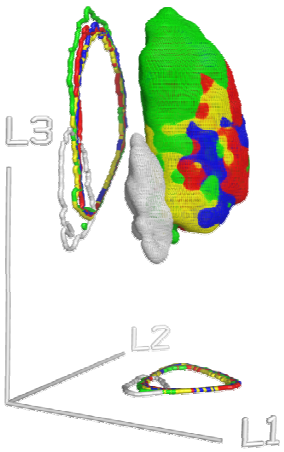
To determine if the different dendrimers exhibit any clear tendencies in terms of shape, and if groups with similar shapes, detached from other shapes, exist (*shape classes*) we employed the same approach we have previously used to study the free energy surfaces constructed using R_g and RMSD. This means that we will identify shape classes in a three-dimensional free energy surface, where the folding coordinates employed correspond to the eigenvalues of **S**. Therefore, we compute the conditional free energy associated with the probability density distribution of **S**

eigenvalues in G_{space} and afterwards we draw 3D surfaces connecting points of equal free energy magnitude in G_{space}.

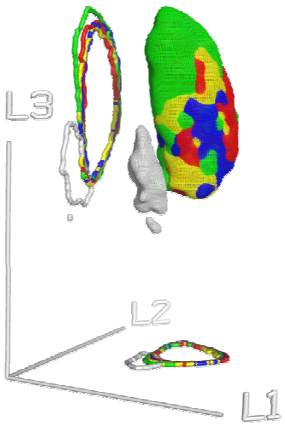
The existence of classes accounting for conformations with shapes more similar among them than with the rest of the conformations (sort of cluster) will be evidenced by the coexistence of surfaces with the same energy value but disconnected (each surface will carry correspondence to a free energy well).



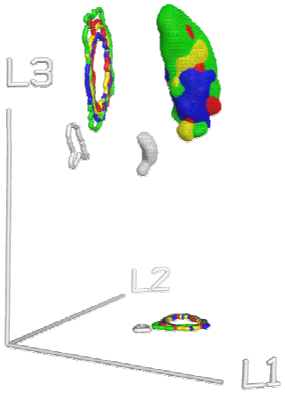
5 RT



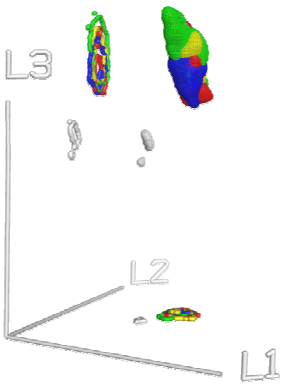
2.5 RT



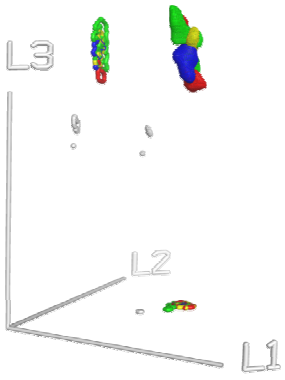
1.5 RT



0.5 RT

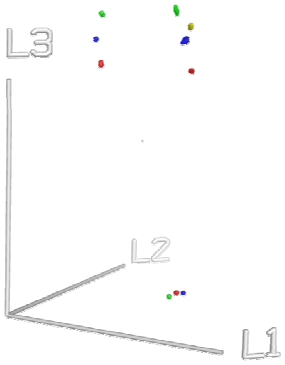


0.25 RT



0.1 RT

Continues on next page



0.01 RT

Figure 3.29 - Free energy isosurfaces for peptide dendrimers in gyration space.

Free energy isosurfaces for peptide dendrimers using as folding coordinates the principal components of the radius of gyration tensor. The surfaces corresponding to different dendrimers are colored accordingly, thus: red for B1, green for B1H, blue for B1HH, yellow for B1HHH and gray for C1. To gain insights into the folding dynamics in this space we have represented the isosurfaces at several energy values (5 RT, 2.5 RT, 1.5 RT, 0.5 RT, 0.25 RT, 0.1 RT and 0.01 RT), that are explicit in the bottom of each 3D-plot. For further details the reader is directed to the text. The contour displayed are 2D projections of the corresponding energy isosurfaces.

In Figure 3.29 we represent the free energy surfaces, and respective contours[43], for all the dendrimers at different free energy values, ranging from 5 to 0.01 RT[44]. With this succession of surfaces at different energy values we hope to clarify the features of the landscape as a function of changes in the shape of the dendrimers.

The figure suggests that, as expected, all B1-family members occupy almost identical regions in terms of the potential shapes they can adopt. This is easily observed from the obvious overlap between the energy surfaces. However, the lowest energy conformations of each of dendrimer occupy different regions of the space, thus presenting some variability in their *native shapes* (see contouor 0.01 RT in Figure 3.30).

[43] The projection of the contour associated with a 3D space, such as the *gyration space*, or other *euclidean* one, can be performed in an infinity of planes. In here, for a matter of schematic coherence, we have projected the energy contours onto the plans defined by (L_1,L_2,0) and (0,L_2,L_3).

[44] We have used a free energy surface of 0.01 RT to pinpoint the conformations closer to zero RT. Otherwise we would have to represent a single point in this huge space making unpractical to visualize it.

Concerning the existence of shape classes we observed two distinct behaviors: B1HH and B1HHH do not present identifiable *classes*; while B1 and B1H exhibit two groups at very low energy values (approximately 0.1 RT). This implies that all B1-family members present folding processes that proceed essentially as downhill ones, where the transitions between conformations (and associated shapes) occur by means of small energy transitions. Furthermore, in B1 and B1H although local minima exist, they are separated from the global minimum by very small energy transitions.

As for C1, it is worth mentioning its relative position is G_{space} once more, in relation to the members of B1-family. C1 appears at a different region of the space, a region featuring lower L_1, L_2 and L_3 values. This is especially evident in Figure 3.29.

The formation of shape classes for C1 is also different from the observed in the B1-family; in C1 we verify the existence of a local minimum at approximately 1.5 RT, this minimum accounts for conformations with a particular shape; as we analyze lower energy values we observe the appearance of yet another local minimum at approximately 0.25 RT. This minimum of energy continues to exist even for very low energy conformations, in the order of 0.1 RT. The identification of the global minimum only becomes clear at 0.01 RT. As a consequence, we must consider that the folding landscape of C1 is somewhat more complex than the B1-family ones, exhibiting at least three *shape classes*. One of these *classes* accounts for conformations with high energy values, and transitions between this class and the others are characterized by high energy barriers. The other two *classes* represent the lowest energy conformations and transitions between these minima are characterized by low energy values. Therefore, it exists a set of C1 shapes varying greatly among them, but still present similar (and low) energy values; and although we have identified an energy minimum we should note that the difference in terms of the energy to the "surrounding" conformations is very small. This idea is further supported by the area occupied by the energy surfaces of C1, which are always smaller than the area of the correspondent surfaces in B1, meaning that the concentration of structures around the existing minima is always higher in C1.

According to what was previously mentioned when we introduced conformational analysis, in dendrimers we can consider the *native structures* as the ones that correspond to the lowest energy. Therefore, using the surfaces drawn through the

principal components of **S**, we can identify these lowest energy conformations for each dendrimer.

In Figure 3.30 we present the structures that correspond to the global energy minimum for each peptide dendrimer, obtained using the free energy surfaces on *gyration space*.

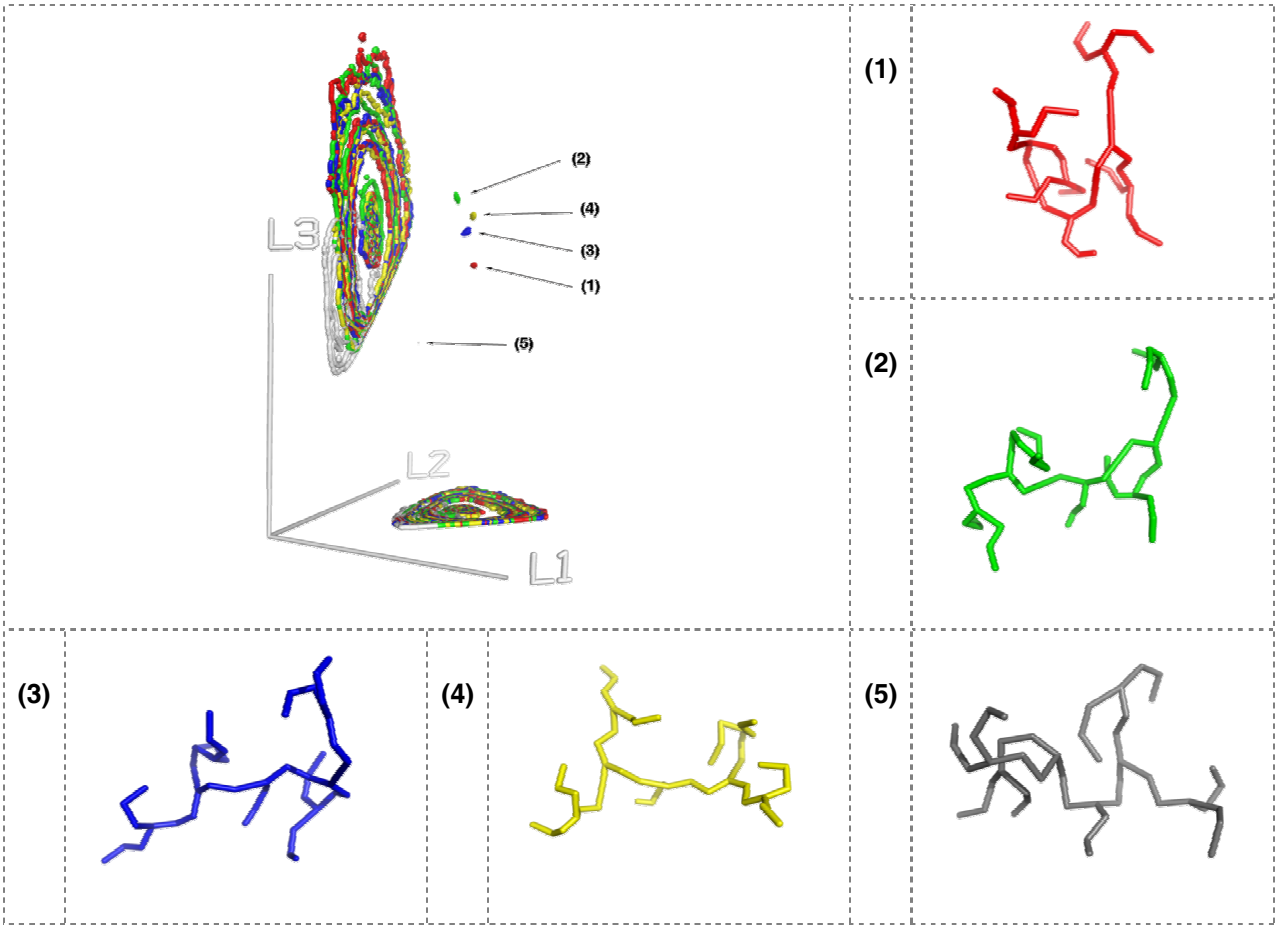


Figure 3.30 – Free energy isosurfaces for peptide dendrimers in gyration space at 0.01 RT.

Free energy isosurfaces for peptide dendrimers using as folding coordinates the principal components of the radius of gyration tensor. The surfaces corresponding to different dendrimers are colored accordingly, thus: red for B1, green for B1H, blue for B1HH, yellow for B1HHH and gray for C1. Alongside with the energy isosurfaces the reader can identify the structures that correspond to the energy minima for each dendrimer. The contours displayed are 2D projections of the energy isosurfaces at several energy values (5 RT, 2.5 RT, 1.5 RT, 0.5 RT, 0.25 RT, 0.1 RT and 0.01 RT). For further details the reader is directed to the text.

From the Figure 3.30, we observe that the native conformations of B1H, B1HH and B1HHH are relatively "stretched", at least when compared to the C1 and B1 ones. If for C1 this kind of open structure with multiple interactions among two of the arms, forming a packed nucleus, was expected, the *native shape* of B1 constitutes a kind of surprise. By analyzing the structure we concluded that, not only it displays significant differences from the ones of the other B1-family members, but it is also a compact structure. Despite not having interlacing branches, it exhibits them near each other and almost align, with the side chains displaying great proximity and interacting. This will be discussed briefly.

Next we present a figure containing a broader set of structures contained, or in the vicinity, of the energy minima observed at 0.1 RT. The intention of this figure is to illustrate the conformational variability within the energetic minima.

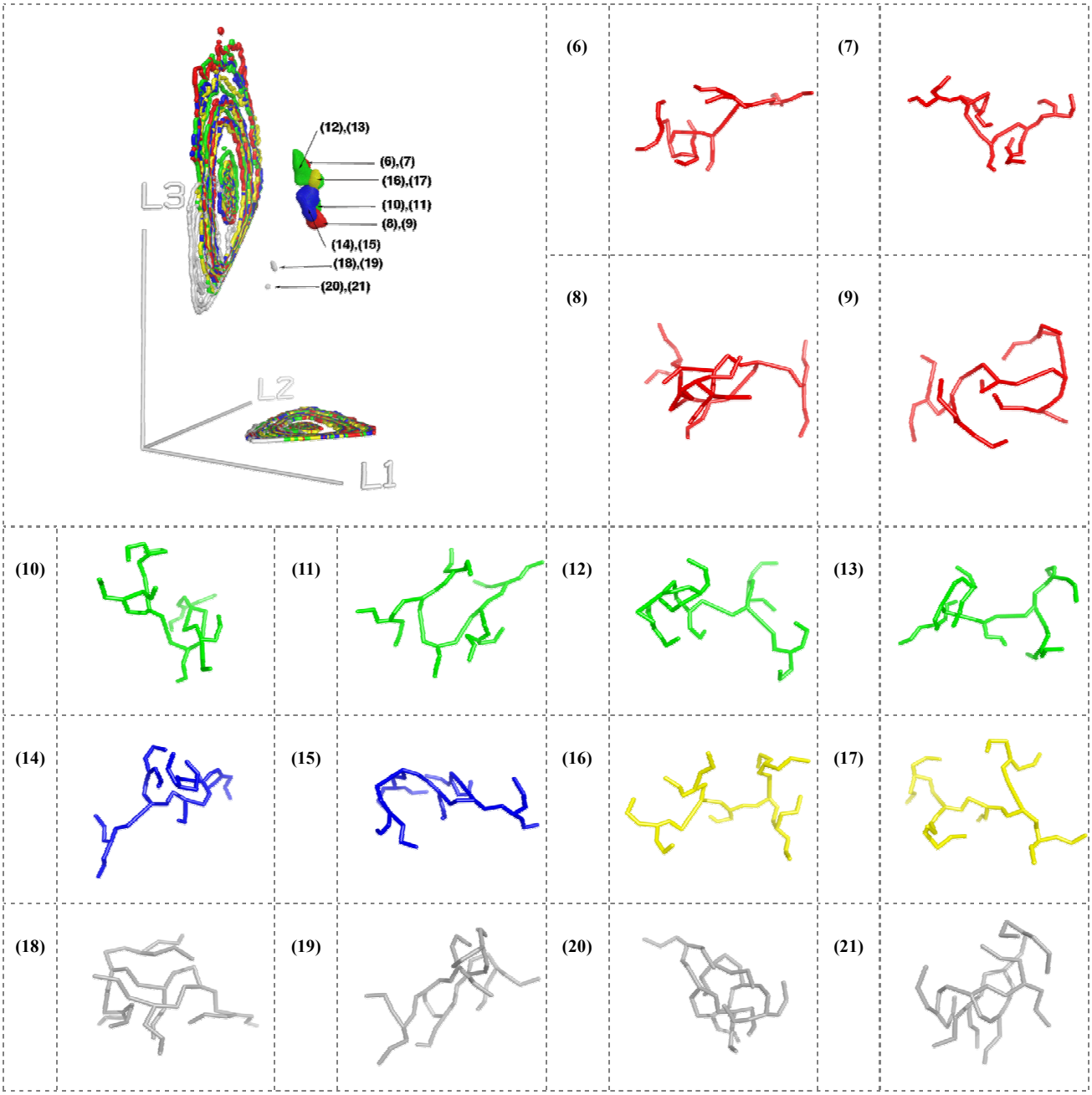


Figure 3.31 – Free energy isosurfaces for peptide dendrimers in gyration space at 0.1 RT.

Free energy isosurfaces for peptide dendrimers using as folding coordinates the principal components of the radius of gyration tensor. The surfaces corresponding to different dendrimers are colored accordingly, thus: red for B1, green for B1H, blue for B1HH, yellow for B1HHH and gray for C1. Alongside with the energy isosurfaces the reader can identify the structures that correspond to the energy minima for each dendrimer. The contours displayed are 2D projections of the energy isosurfaces at several energy values (5 RT, 2.5 RT, 1.5 RT, 0.5 RT, 0.25 RT, 0.1 RT and 0.01 RT). For further details the reader is directed to the text.

As can be verified from a joint analysis of Figures 3.30 and 3.31, the atomic geometries present within each energy basin are variable, and conformations that seem to be structurally different present themselves close in *gyration space*. Which, by looking at Figures 3.27 and 3.28, implies that their asphericity and acylindricity values are also very close.

Still, even though the B1 and C1 lowest energy minimum (represented at 0.01 RT) are different from what expected, if we make just small variation on the value at which we draw the energy isosurface (as represented in Figure 3.30), we immediately observe the "expected" atomic geometries.

The type of approach used here has proven to satisfactorily describe the shape of dendrimers, but another question arises: is the analysis of the shape, namely the use of gyration space, also a good approach to perform conformational analysis? In other words, is *gyration space* also a suitable *conformational space*?[45]

To answer this question we introduce Figure 3.32, that renders the distance between all pairs of points in *gyration space*[46], representing the (dis)similarity between the corresponding structures in that space, and one the most common (dis)similarity measures, the RMSD between all pairs of conformations.

To accomplish this we must take into account the computational cost involved in the determination of the RMSD matrices for the entire set of conformations[47]. Moreover, to achieve the desired accuracy we must compute 128 RMSD matrices, accounting for all the possible permutations (see section 3.2.7.). Using the ensemble of structures considered for the entire shape analysis is impossible, thus we have chosen a representative set of conformations composed by structures extracted at intervals of 1 ns from each replicate equilibrated trajectory. This accounts for a total

---

[45] A suitable conformational space, in the case of dendrimers, would be a *n*-dimensional space capable of accurately discriminate conformations with different structural features.

[46] The distance between two points in a three-dimensional space is defined as the norm of the vector defined by them. In this particular case we calculate it as: $d_{ij} = \sqrt{\left(L_1^i - L_1^j\right)^2 + \left(L_2^i - L_2^j\right)^2 + \left(L_3^i - L_3^j\right)^2}$, where *i* and *j* represent different points in G_{space} and d_{ij} the distance among them.

[47] If we consider an ensemble composed by *N* conformations, the corresponding RMSD matrix accounting for the (dis)similarity between all pairs of structures will have N^2 entries.

of 860 (B1), 915 (B1H), 915 (B1HH), 890 (B1HHH) and 845 (C1) reference structures, each with its own 128 permuted conformations.

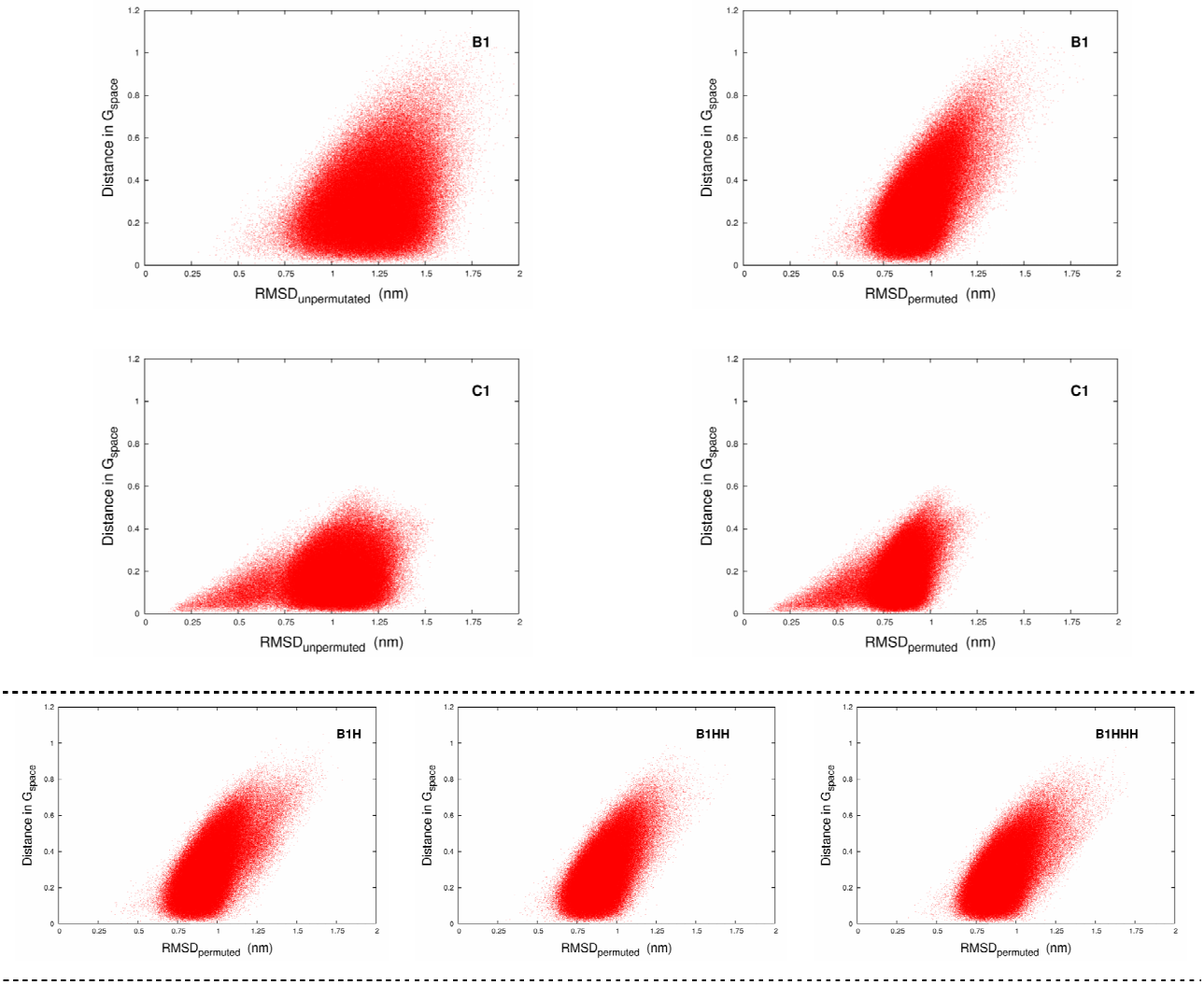


Figure 3.32 – Gyration space as a conformational space.

Distance among all pairs of structures in gyration space, as a function of $RMSD_{permuted}$ and $RMSD_{unpermuted}$. See the text for details.

The scatter-plots presented in the previous figure evidence a satisfactory conformational discrimination by G_{space}, that is however, far from being perfect, as there still exist conformations with identical distances in G_{space} but distinct RMSD values. This discrepancy seems to be most obvious for C1. Apart from the hypothesis that the amino acid residue composition promotes this difference (which is excluded by looking into the scatter-plots for B1 and B1HHH, where dendrimers with a

somewhat different amino acid residue composition present an almost identical point distribution), a possible explanation is that the higher symmetry presented by C1, evidenced by the shape analysis, is probably conferred by the ubiquitous presence of Dap as a branching residue, and is responsible by the greater conformational "inexactness" of its G_{space} (when compared to the other dendrimers under study).

Next, we present Figure 3.33, where we plotted the values of *c* and *b* as a function of R_g, for each of B1 and C1 conformations previously represented in the G_{space} (see Figures 3.27 and 3.28). This helps to elucidate the correlation between the different properties/measures.

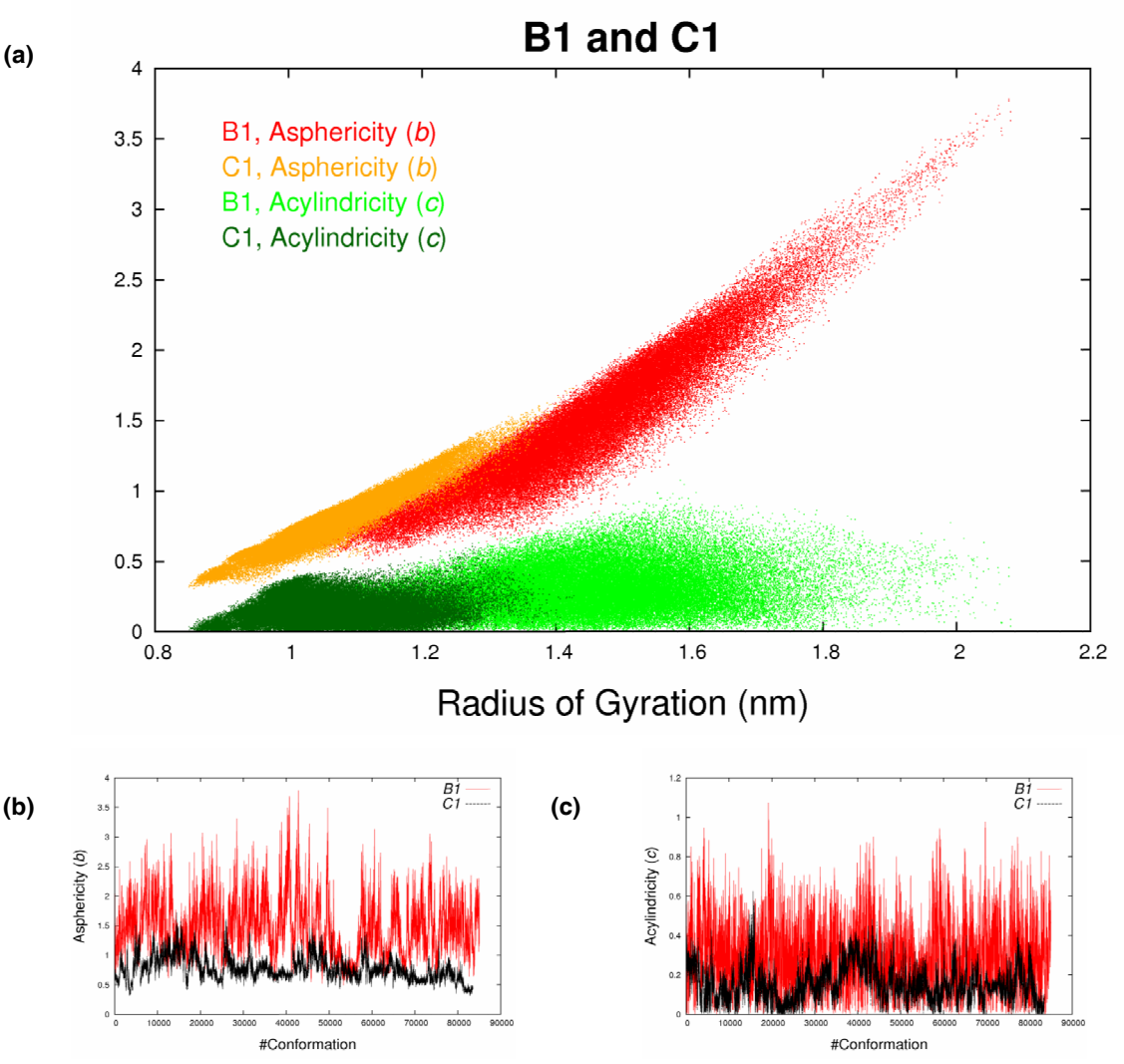


Figure 3.33 – Relation between shape descriptors and the radius of gyration.

(a) Plot combining the values of c and b for B1 and C1, in function of Rg. (b) Values of asphericity for all the equilibrated conformations of B1 and C1. (c) Values of acylindricity for all the equilibrated conformations of B1 and C1. See the text for details.

Analyzing the figure, we conclude that all shape descriptors share similar trends towards R_g (in terms of the shape of the data distribution), which is not strange, since they are all calculated based on the components of the diagonalized matrix of the R_g tensor. Still, the values of C1 are confined to a smaller area of the plot then the ones of B1.

Examining the plots where we represent *c* and *b* as a function of the a randomly assigned conformation number[48] it becomes clear that, for the set of conformations considered, B1 presents higher *c* and *b* values than C1. This implies that B1 has, globally, less symmetrical conformations than C1. Furthermore, B1 has also a less pronounced cylindrical symmetry than C1.

To confirm these findings we have determined average values for *c* and *b* in the aforementioned ensembles.

Table 3.6 – Average values of the shape descriptors for B1, B1H, B1HH, B1HHH and C1.

Dendrimer	$\langle b \rangle$	$\langle c \rangle$
B1	1.4803	0.3019
B1H	1.5176	0.2937
B1HH	1.4535	0.3102
B1HHH	1.4292	0.3022
C1	0.7786	0.1547

[48] The conformation number we refer to is a random one, because as the conformations are obtained from different replicates they do not possess an intrinsic organization. What is important to observe is the magnitude of the properties.

As expected B1, B1H, B1HH and B1HHH exhibit similar values for asphericity and acylindricity. On the other hand, C1 presents average values for *c* and *b*, that are almost half the ones determined for the other dendrimers. From examining Figure 3.3

3 it is also clear that acylindricity is the shape descriptor less correlated with the radius of gyration (what makes sense, because *c* is calculated taking into account just L1 and L2). Using the acylindricity values as a reaction coordinate along with $R_{g,}$ we were able to build the corresponding energy surfaces (2D). The folding dynamics, as well as the conformational minima are similar to the ones obtained for the energy surfaces of gyration space (data not shown).

Given all that was mentioned about dendrimers shape, and its relation with their conformational analysis, we must conclude that even in terms of the ellipsoid accounting for the shape of dendrimers we observe that the different dendrimers, can present a myriad of possible disparate shapes. Maybe the most important and unexpected (according to the experimental suggestions [Sommer 2008]) conclusion to be drawn from this section is that the B1-family dendrimers do not resemble spheres. In fact, they present highly disordered structures with multiple conformational and shape possibilities.

3.2.10. Principal Coordinate Analysis

To complement the information supplied by histogram, free energy and shape analysis, we resort to a methodology vastly applied in different areas of knowledge (from advertising to biochemistry): *metric multidimensional scaling* or *principal coordinate analysis* (*PCoorA*) [Cox 2001].

We propose an approach for summarizing the distribution of conformations, that acknowledges the complexity and structural flexibility intrinsic to the dendritic system we are studying, and promotes a low-dimensional representation of the configuration space explored by the different systems.

First, we compute a dissimilarity matrix for all pairs belonging to an ensemble of conformations, and employ that information in conjunction with PCoorA [Becker 1996; Hamprecht 2001; Rencher 2002] to obtain a low-dimensional representation of the conformational space. After that, a nonparametric estimate of the density of states in this subspace is obtained using kernel methods. The corresponding free energy surface is then calculated from that density, and the ensemble configurations are clustered in such a way that all conformations belonging to one local free energy minimum form a conformational *class*.

The free energy values were obtained for three dimensions and can be directly visualized using energy isosurfaces. Our approach will use a particular dissimilarity measure, RMSD, computed for all pairs of conformations in an ensemble and embed each conformation as a point in the configuration space.

As previously mentioned, creating useful representations of molecular conformations space is a hard task (see section 3.2.1.), hindered by the high dimensionality of these spaces. Therefore, collective coordinate methods, such as *principal components analysis*, *normal mode analysis* or PCoorA, allow the projection of multidimensional data on low-dimensional subspaces, and are thereby very suited for representing and visualizing conformation spaces and MD trajectories that transverse these spaces.

The goal of PCoorA is reducing the number of dimensions that describe a complex system from its complete set to a lower-dimensional representation that contains the most important data characterizing that system [Steinhauser 2005].

In the present work, the data to be retained are the generalized coordinates that distinguish conformations of dendrimers, moreover, these should be coordinates that show the essential distinction between structures that lie far from each other on their many-dimensional potential surface [Steinhauser 2005].

There are a few examples of the use of PCoorA in the study of molecular processes and conformations, mostly related to the analysis of the potential energy surfaces of proteins [Becker 1996; Hamprecht 2001; Levy 2001(a); Levy 2001(b)].

A detailed description on PCoorA is out of the scope of this thesis. Still, some brief considerations are presented in Appendix H. For further information on principal coordinate analysis and other collective coordinate methods the reader is directed elsewhere [Cox 2001; Hamprecht 2001; Rencher 2002].

As referred in Appendix H, PCoorA requires the diagonalisation of a $n \times n$ matrix, where n stands for the number of conformations of the system. Thus, with the currently available computational power, this procedure is unfeasible for large datasets (more than a few thousand conformations). Herein we will use an RMSD matrix constructed from 4251 (B1), 4526 (B1H and B1HH), 4401 (B1HHH) and 4176 (C1) configurations extracted from the concatenated trajectories at intervals of 0.2 ns. The RMSD matrices were build using the permutation methodology previously described (see section 3.2.7.). The program used to compute the PCoorA output was *mds* (kindly provided by Paulo J. Martel; Universidade do Algarve).

We chose to reduce the dimensionality of the data set to a three-dimensional space. This choice has to do with the fact that we want to retain as much information as possible but still be able to easily visualize our results.

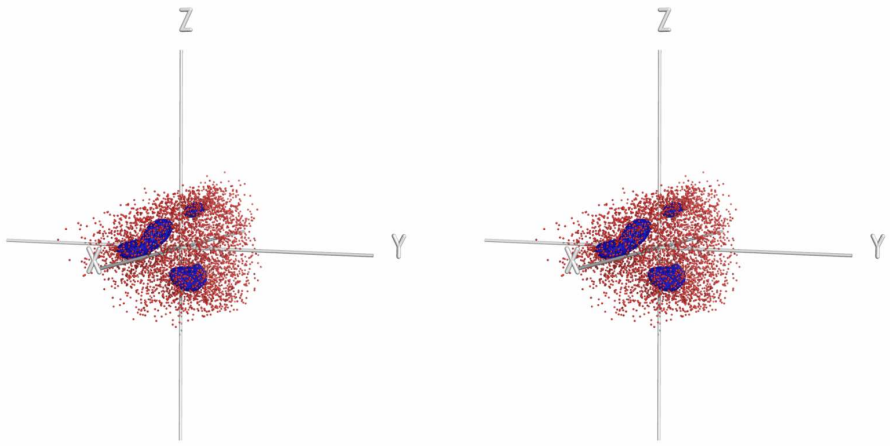
Once more, with the use of free energy surfaces (see section 3.2.7) we propose to cluster the configurations using as "membership" criterion the presence (or absence) within a free energy minima, or, equivalently, local density maxima [Campos 2009]. With this approach we expect to identify what van Gunsteren and coworkers defined as "islands of stability" in the configuration space [Hamprecht 2001].

3. Results and Discussion

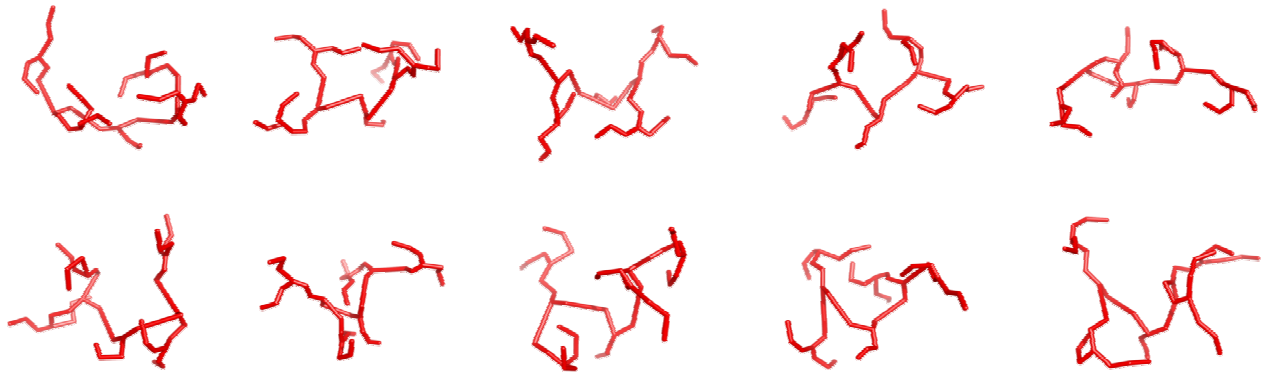
The results obtained for each dendrimer, employing PCoorA in a 3D space, while using RMSD as dissimilarity (or distance) measure are presented next.

We chose the presented RT values because they are the ones exhibiting the higher number of clusters.

0.1 RT

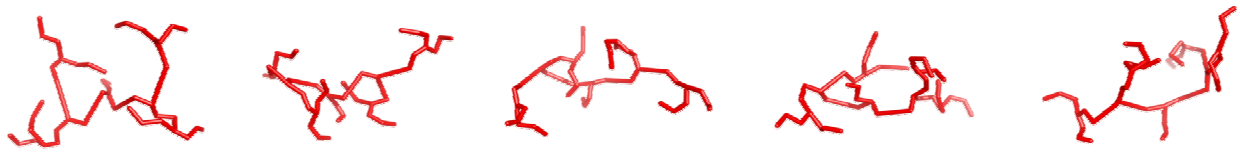


(1)



1.86%

(2)



2.05%

Continues on next page

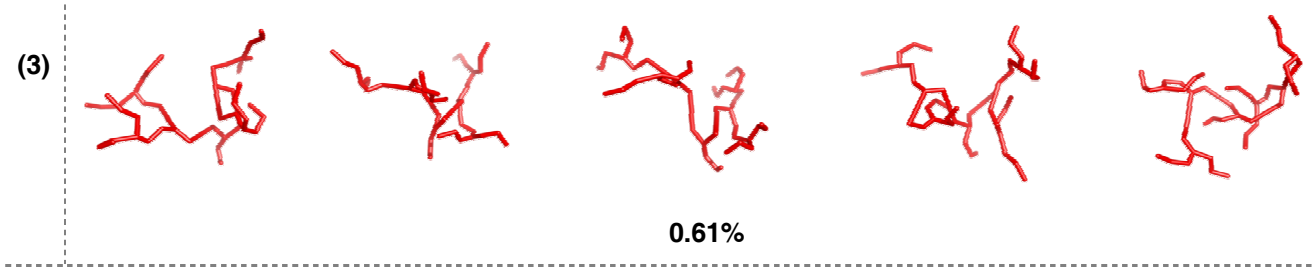


Figure 3.34– B1 principal coordinate analysis results and associated energy isosurfaces.

To get the most of the 3D image please use cross-eye stereo. The red dots represent the results obtained by applying 3D-PCoorA to a matrix containing the permuted RMSD values between all conformations in a B1 ensemble, with 4251 configurations.

The energy isosurfaces were drawn at 0.1 RT. The examples of conformations associated with each of the three isosurfaces are represented as **(1)**, **(2)** and **(3)**; where **(1)** corresponds to the energy surface accounting for the lowest energy conformations, in particular the global energy minimum. The lowest energy conformation is represented under **(1)**, and is the first one on the left. The conformations pinpointed under **(2)** exemplify structures for the second lowest energy cluster. The structures from **(3)** correspond to a cluster with higher energy values than **(1)** or **(2)**. Still the energy differences among the different clusters are small. The values of the relative percentage within each cluster are displayed at the bottom of the conformation examples. See further details on the text.

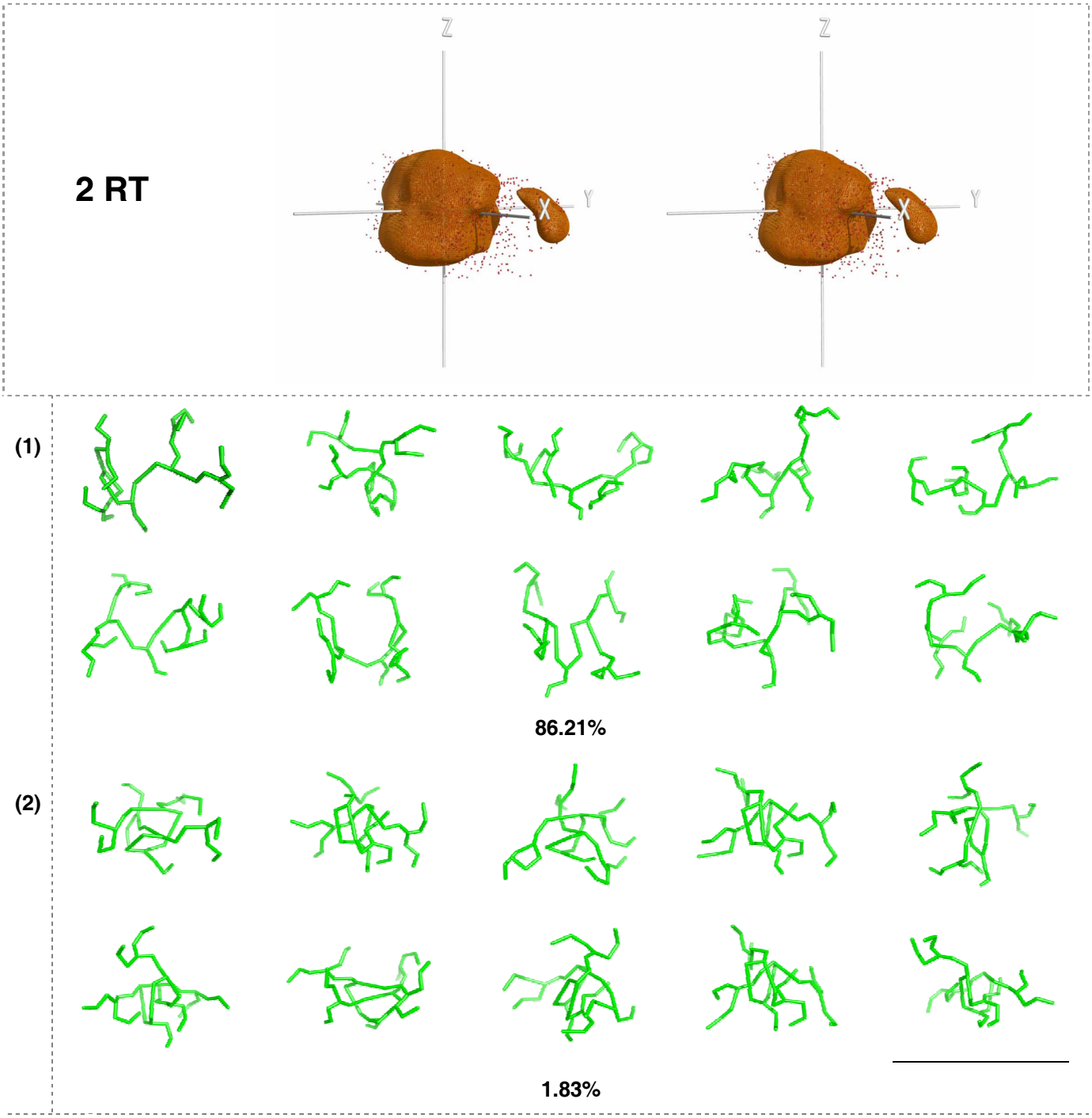


Figure 3.35– B1H principal coordinate analysis results and associated energy isosurfaces.

To get the most of the 3D image please use cross-eye stereo. The red dots represent the results obtained by applying 3D-PCoorA to a matrix containing the permuted RMSD values between all conformations in a B1H ensemble, with 4526 configurations.
The energy isosurfaces were drawn at 2 RT. The examples of conformations associated with each of the two isosurfaces are represented as **(1)** and **(2)**; where **(1)** corresponds to the energy surface accounting for the lowest energy conformations, in particular the global energy minimum. The lowest energy conformation is represented under **(1)**, and is the first one on the left. The conformations pinpointed under **(2)** exemplify structures for the second energy cluster. The values of the relative percentage within each cluster are displayed at the bottom of the conformation examples. See further details on the text.

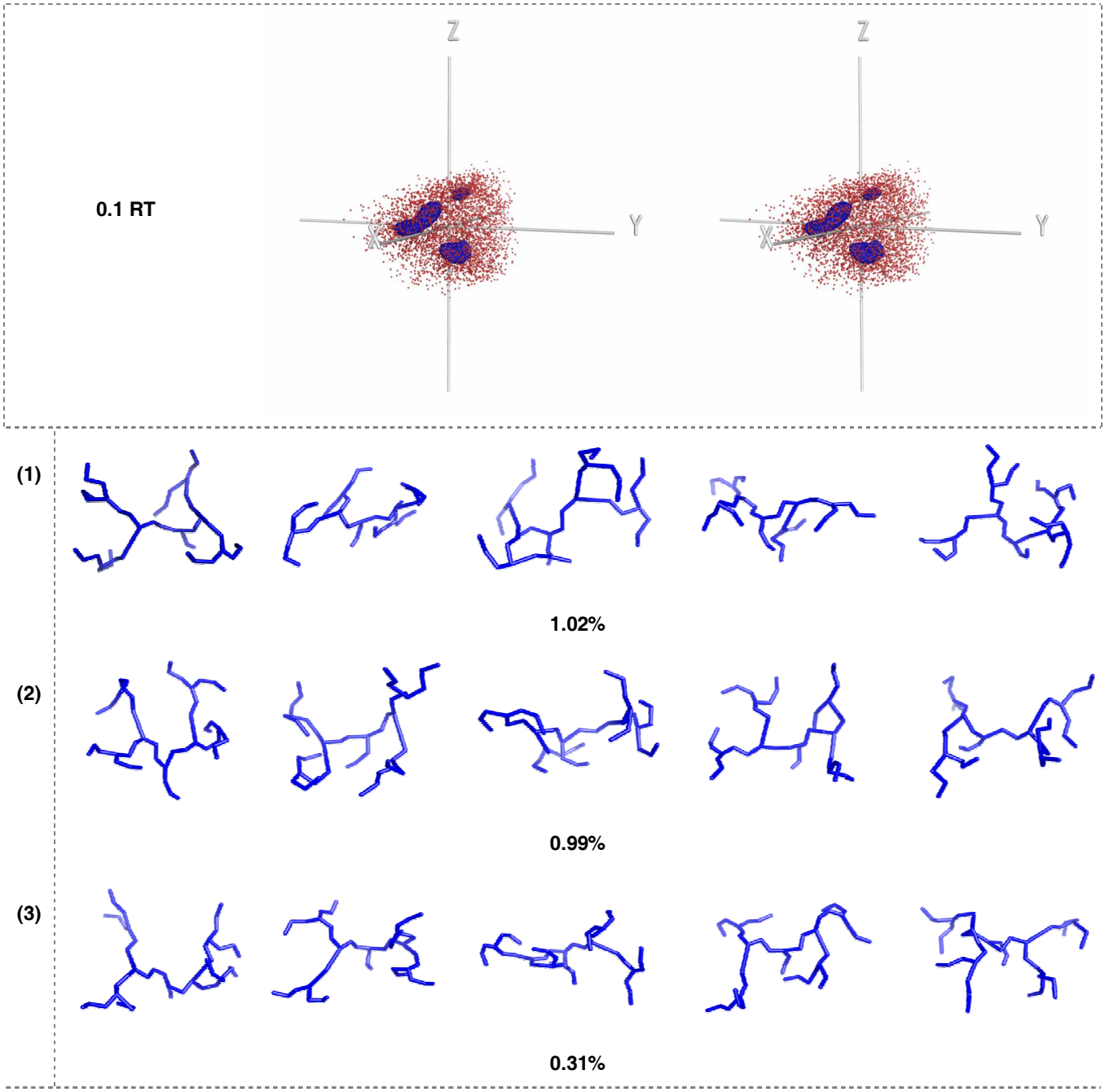


Figure 3.36– B1HH principal coordinate analysis results and associated energy isosurfaces.

To get the most of the 3D image please use cross-eye stereo. The red dots represent the results obtained by applying 3D-PCoorA to a matrix containing the permuted RMSD values between all conformations in a B1H ensemble, with 4526 configurations.
The energy isosurfaces were drawn at 0.1 RT. The examples of conformations associated with each of the three isosurfaces are represented as (1), (2) and (3); where (1) corresponds to the energy surface accounting for the lowest energy conformations, in particular the global energy minimum. The lowest energy conformation is represented under (1), and is the first one on the left. The conformations pinpointed under (2) exemplify structures for the second lowest energy cluster. The structures from (3) correspond to a cluster with higher energy values than (1) or (2). Still the energy differences among the different clusters are small. The values of the relative percentage within each cluster are displayed at the bottom of the conformation examples. See further details on the text.

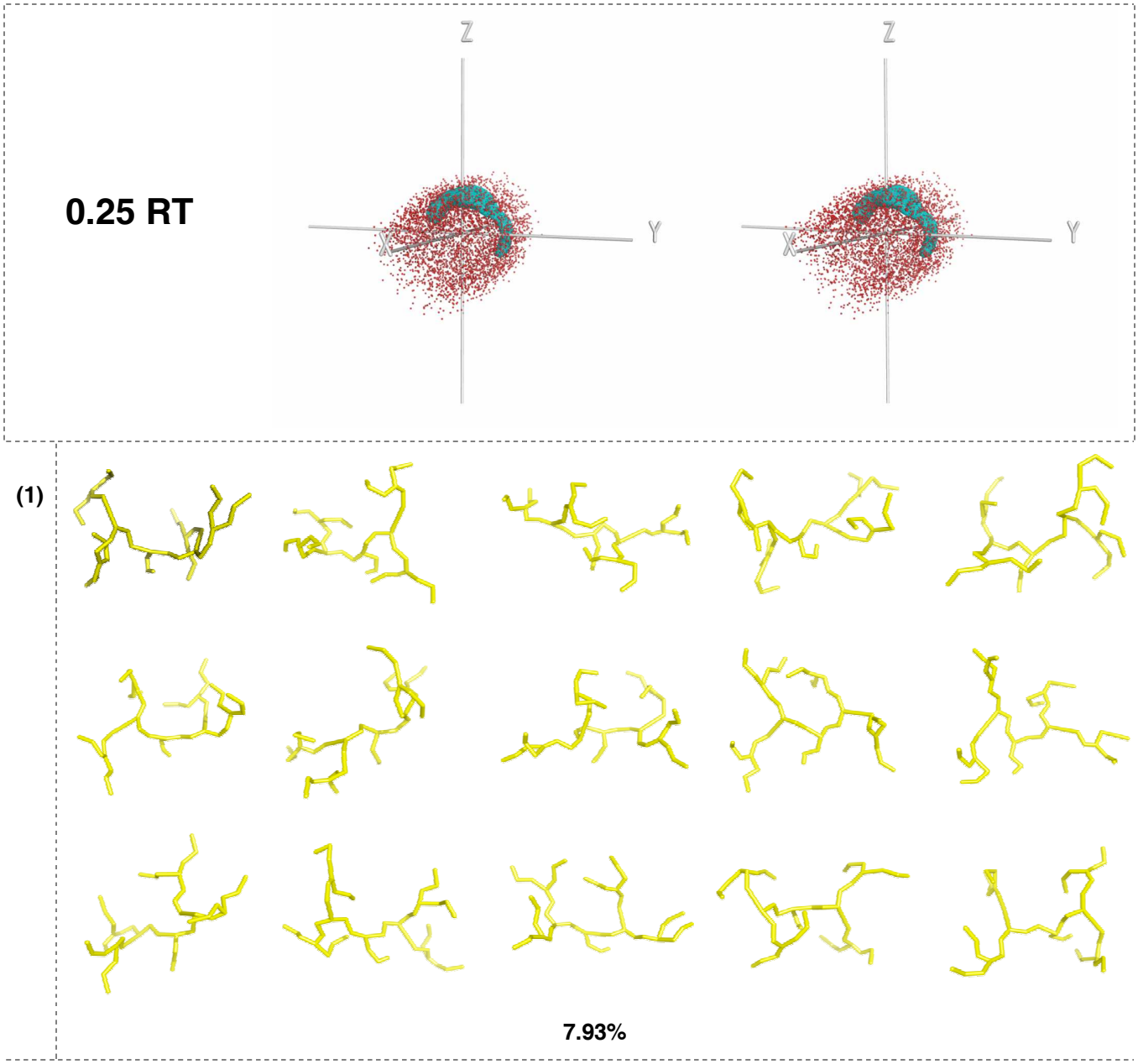
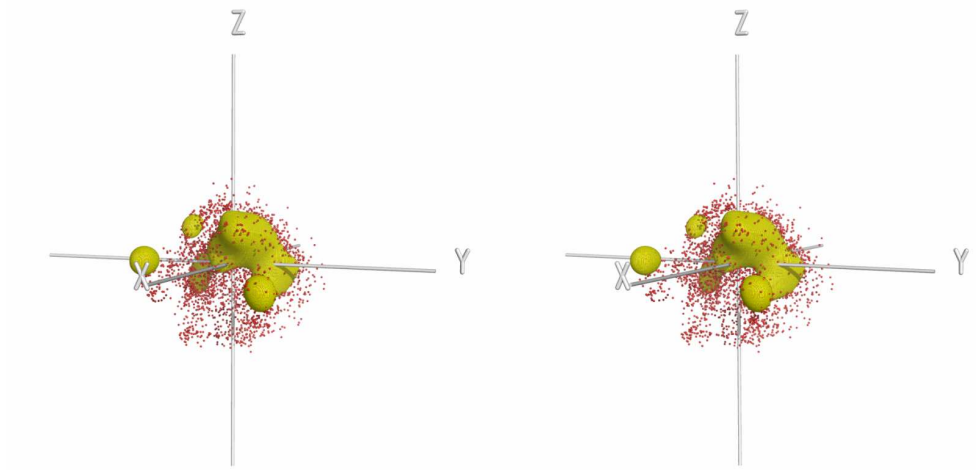


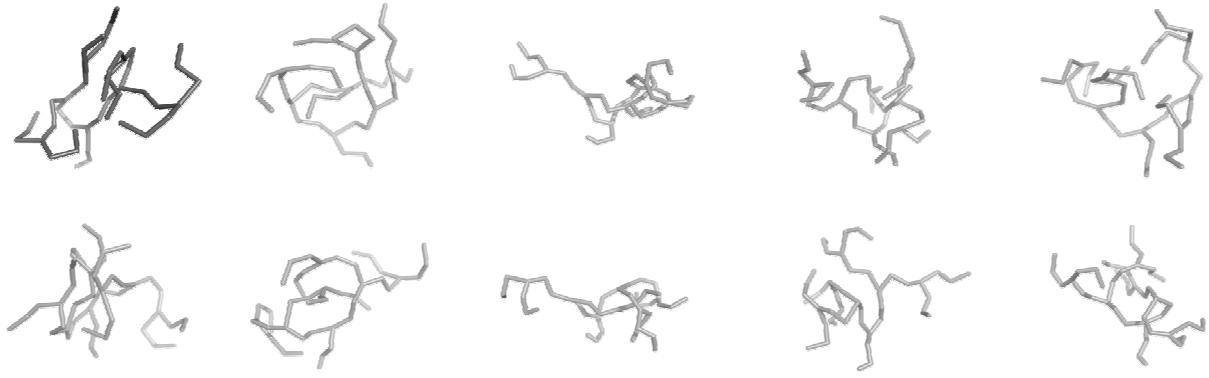
Figure 3.37– B1HHH principal coordinate analysis results and associated energy isosurfaces.

To get the most of the 3D image please use cross-eye stereo. The red dots represent the results obtained by applying 3D-PCoorA to a matrix containing the permuted RMSD values between all conformations in a B1HHH ensemble, with 4401 configurations.

The energy isosurface were drawn at 0.25 RT. The examples of conformations associated with the isosurface are represented as **(1)**; The lowest energy conformation is represented under **(1)**, and is the first one on the left. The values of the relative percentage within the cluster are displayed at the bottom of the conformation examples. See further details on the text.

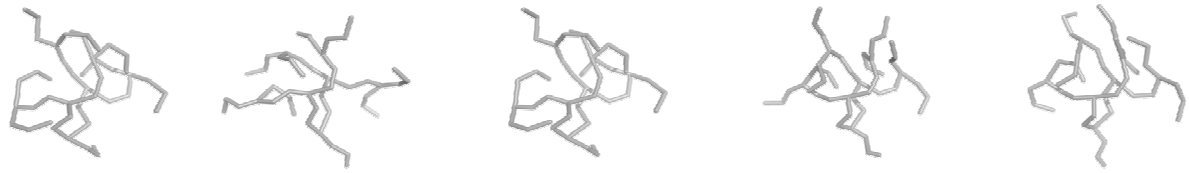


(1)



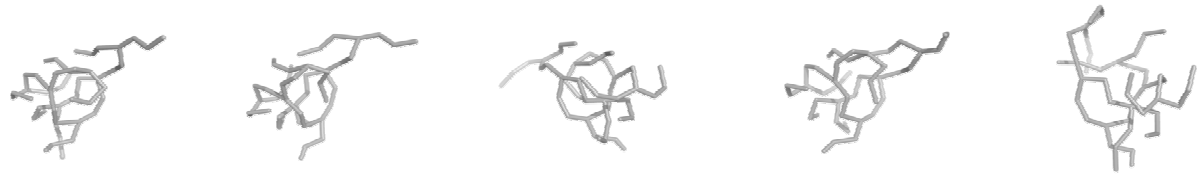
19.32%

(2)



4.41%

(3)



3.88%

Continues on next page

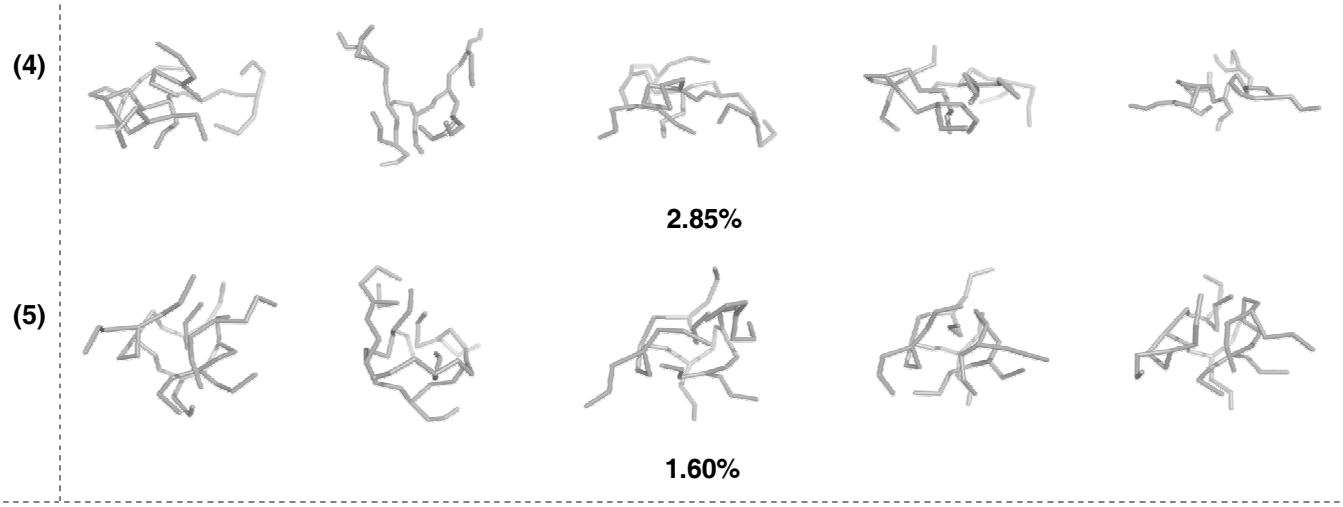


Figure 3.38– C1 principal coordinate analysis results and associated energy isosurfaces.

To get the most of the 3D image please use cross-eye stereo. The red dots represent the results obtained by applying 3D-PCoorA to a matrix containing the permuted RMSD values between all conformations in a C1 ensemble, with 4176 configurations.

The energy isosurfaces were drawn at 1 RT. The examples of conformations associated with each of the three isosurfaces are represented as **(1)**, **(2)**, **(3)**, **(4)** and **(5)**; where **(1)** corresponds to the energy surface accounting for the lowest energy conformations, in particular the global energy minimum. The lowest energy conformation is represented under **(1)**, and is the first one on the left. The conformations pinpointed under **(2)**, **(3)**, **(4)** and **(5)** exemplify structures belonging to cluster with different energy values; the numeration was assigned in an ascending gradient, according to the energy with which the different clusters "disappear" on the global energy landscape. The values of the relative percentage within each cluster are displayed at the bottom of the conformation examples. See further details on the text.

The detailed features of the landscape are not accessible from the 3D plots presented. Therefore along the text we will mention the events occurring for different contour values of free energy, thus trying to conceptualize the main characteristics of each hypersurface.

The first obvious result is that B1 and B1HH present similar (almost identical) "islands of stability", positioned almost in the same regions and covering similar amounts of conformations (in terms of percentage).

For B1 the only observable clusters appear at 0.1 RT, a very low value of energy that separates energy minima accounting for heterogeneous but energetically close

conformations. Despite the heterogeneity, all conformations present common trends, i.e., "mildly stretched" structures with few interactions among the different branches. Furthermore, the use of slightly different contour values (not shown) indicates that the transitions between the different energy minima are characterized by very small free energy barriers and thus, these conformations can be considered as kinetically very close in B1 conformational space.

Moreover, since this is the only energy value, in the entire energy scale, at which we observe "islands of stability", one can consider that the folding process for B1 using this folding coordinates is a downhill process, characterized by a flatten surface with a rough peak.

The folding landscape for B1HH reflects a folding dynamics similar to B1, where even the conformations represented in the low energy minima share the same structural features: "mildly stretched" structures with few interactions among the different branches.

B1HHH does not presents any free energy clusters (or local minima), in fact the isosurface represented in the figure (and that corresponds to 0.25 RT), simply accounts for a value of free energy, chosen to represent the position of the lowest energy conformations. It is therefore a clear case of downhill folding towards a single global minimum.

Once more we verify that the lowest energy conformations are very similar, with the dendritic branches widely distributed in *bowl-like* conformations.

Thus, for the aforementioned dendrimers (B1, B1HH and B1HHH) we cannot consider the existence of "true islands of stability" in this conformational space. Moreover, the low energy clusters of B1 and B1HH correspond to very small sets of the entire population (approximately 1.5 %) and can hardly be considered as representative.

It is also important to note that the *folding behaviors* of B1, B1HH and B1HHH observed in this space are very similar to the ones obtained using other reaction coordinates, namely RMSD and R_g (2D) or the principal components of the radius of gyration tensor (3D).

For B1H it is possible to observe the formation of two distinct conformational groups, where the smaller one constitutes a clearly distinct energy well (local

minimum) between 2.5 RT and 1.25 RT, after those values the corresponding isosurface vanishes. The main group "proceeds" into the global minimum without the appearance of any other "island of stability".

In fact, the local energy minimum of B1H accounts for a small number of high energy structures (1.83%) that, despite being conformationally heterogeneous, generally reflect sphere-like (or molten-globule like) conformations, where the residues are packed together and some contacts between the different branches are observable. In effect, for B1H, as for all the B1-family dendrimers, the more compact and packed structures are not the ones that minimize the overall free energy, and although they always exist (see also Figure 3.11 they correspond to small percentages of the total conformational population. Still, these conformations might explain the existence of a small frequency peak, for low R_g values (in the order of 1.05 nm) in the histogram for B1H (see Figure 3.7), and that had not been satisfactorily discriminated by the conformational analysis performed until now. The energy barriers between this local minimum and the global one, are high energy ones (approximately 1.25 RT).

The main isosurface of B1H represents almost all the conformations (approximately 87 %) that do not belong to the smaller isosurface, gathering heterogeneous conformations that seem to display a similar trend, and are generally more loose than the ones from the local minimum.

The folding dynamics of B1H seems to be discriminated differently when using different approaches. With PCoorA the conformational space evidences the existence of a small set of high energy conformation, that were not as patently separated in the folding landscapes obtained employing other folding coordinates (see Figure 3.23and Figure 3.29.

For C1 the situation differs markedly from the ones previously presented for the other dendrimers. In this case, we can distinguish different free energy clusters, starting at 3 RT and raging to very low energy values (0.1 RT except for the one that corresponds to the global energy minimum, that goes to zero). Therefore, we verify the existence of "islands of stability" for C1 conformations. Some of the different isosurfaces account for representative percentages of the total amount of conformations. As the values for the energy barriers between the different mimima

vary, we can consider that the folding landscape of C1 is a complex and rough one, presenting a few local minima disposed beneath 3 RT.

We can successfully identify some clear C1 conformational trends, and some structural features characterizing each energy cluster. This is evident from looking at the configurations randomly collected from each isosurface. Almost all free energy isosurfaces account for structures with packed areas and multiple interactions between the constituting residues. As a consequence, we must conclude that the conformational space created using the folding coordinates provided by PCoorA, is the one that more efficiently discriminates the different conformational preferences of C1 (that definitely exist). Moreover, the conformations present at the conformational minimum are structurally equivalent to the ones obtained using other reaction coordinates; and in fact the preferential structures that C1 "populates" are compact ones (more compact and with more inter-residue contacts than any of the other peptide dendrimers studied here). Still, it does not favor the most compact/packed ones, although it can visit them (this was to be expected, see Figure 3.11).

As a concluding remark we must emphasize that PCoorA seems to be the conformational space more suited (from the ones used in these thesis) for studying the folding of peptide dendrimers. Because it not only reveals the features identified using other methodologies (such as shape analysis, R_g vs. RMSD, histogram analysis), but also allows for the identification of new ones, that were "hidden" in the complexity of the data (e.g., the high energy group in B1H).

Chapter 4

Concluding Remarks

4.1. Conclusions

The first noteworthy conclusion (and maybe the principal one) of this study is that, contrarily to what as been advocated by other authors [Darbre 2006;Javor 2009; Sommer 2009], our simulations suggest the existence of at least two very distinct preferential conformations that peptide dendrimers can undergo. The first type corresponds to compact configurations with multiple interacting branches (that some might consider similar to spheres or molten globules, although in our simulations they do not present at any time a perfect spherical symmetry, see section 3.2.9 on shape analysis). The other type of conformations corresponds to a wide open molecular configuration, where the branches are disposed in a spaced manner, and where the points of contact, other than the bonded ones, are scarce, yielding conformations similar to bowls.

This conclusion alludes to the fact that peptide dendrimers, depending on their generation and amino acid residue composition, may in fact constitute a class of synthetic molecules with the potential to exhibit a myriad of conformational behaviors. Thus, exploring this diversity might be useful in many areas of knowledge.

In fact, we have observed that in some peptide dendrimers (such as B1, B1H, B1HH and B1HHH) the favored atomic geometries are not compact at all, but rather bowl-shaped; at least in terms of the relative accessible states. Some more compact states do exist (not as compact as the ones in C1) but they represent only a very small fraction of the total amount of accessible conformations (see sections 3.2.8 and 3.2.10).

These dendrimers exhibit folding landscapes (along different sets of reaction coordinates) that account for a process that is essentially a downhill one, towards a single conformational minimum.

These results spread the discussion on the behavior of peptide dendrimers in solution, because they counteract the hypothesis proposed by Reymond and coworkers [Sommer 2008], that both B1 and C1 were not "organized" enough to bind the corrin ring as tightly as natural proteins. In fact, C1 seems to visit states that are preferentially organized with high inter-residue contacts, and shapes resembling molten globules (or sphere-like). Its dynamics is represented by complex folding landscapes, where some conformational energy minima are present, but they all account for "closed" conformations. In that manner C1 is, in principle, organized enough to coordinate aquocobalamin by completely secluding it from solution. Our results however, point toward another direction.

The results presented along this thesis suggest that the dendrimers with higher experimental coordination to aquocobalamin, B1 and B1H (see Table 3.1), preferentially adopt flexible and loose (open or bowl-like) shapes, characterize by few inter-residue contacts. This suggests that peptide dendrimers with high core exposure (and hence, high coordination residue exposure, either Cys or His) have a higher potential to coordinate aquocobalamin.

If we take another look at the experimental values obtained for the hydrodynamic radius of B1 while taking into account our findings (see Table 3.1), we verify that, in light of this new "conformational context", the small differences observed between the hydrodynamic radius of the free form (R_H = 1.840 nm) and the complexed form (R_H = 1.838 nm), are a reflection of the flexibility exhibit by B1. In other words, either in coordination or floating freely in solution, B1 probably exists as an open

structure with its core exposed, and that is probably "the secret behind its success" for coordinating aquocobalamin. Whereas C1 presents the coordinating histidine residues in a more advanced position (see Table 1.2), and as such these residues are probably involved in the formation of the "characteristic" compact structures, consequently becoming more inaccessible to aquocobalamin. This hypothesis would explain the apparent dichotomy of peptide dendrimers with a higher number of coordinating residues (C1 has two histidines) displaying lower affinity towards aquocobalamin, than peptide dendrimers with less coordinating resides (case of B1 and B1H that only have one coordinating residue, cysteine and histidine respectively).

A noteworthy observation is that despite having almost the same amino acid residue composition and conformational preferences, B1 displays an affinity to aquocobalamin that is sixty times higher than the one exhibit by B1H (see 1.3.), where the only parameter that changes is the coordinating residue. Suggesting that for aquocobalamin, cysteine is a more efficient (better if you will) coordination residue than histidine.

Until now, we have only mentioned the exposure of coordinating residues as a factor conditioning the potential for coordination, but the disposition of the surrounding branches must not be ignored. B1 and B1H present, in most of their conformations, widely spaced branches (with regard to the core), and hence the branches might even assist the exposed coordination residue, in the coordination, by stabilizing the aquocobalamin (through the bowl-like shape). In C1 the branches (that also contain the coordinating histidine residues) are involved in a more compact structure where they probably become less available to "aid" in the coordination.

Based on our model for the interaction of aquocobalamin with B1-family peptide dendrimers, we would expect, that if B1HH had been synthesized, it would probably present itself as the dendrimer with highest affinity towards aquocobalamin; as it presents conformational preferences similar to B1, but features three coordination residues. Furthermore, in B1HH the residues replaced by histidines were Amb, which due to their atomic structures and lack of side chain are probably not involved the potential stabilization of aquocobalamin by the dendritic branches.

The structural flexibility and miscellaneous conformational behavior of peptide dendrimers, evidenced in the present work, leads us to consider an analogy with

proteins (once more). Different peptide dendrimers are most likely classified under *dendrimer families*, in attention to their favored conformations. Dendrimers that preferentially assume a compact structure (such as C1 and the ones observed by Reymond and coworkers [Javor 2009]) will form a family, whereas dendrimers with open, bowl-like configurations will form another (such as the ones in the B1-family[49]). Additional peptide dendrimers exhibiting structural behaviors not accounted in the existing families, will form new ones; as with proteins. Still, this is just a divagation and a potential glance at the future.

An important question raised by the present results concerns the factor(s) determining the shape of peptide dendrimers. Which is the most influential factor(s) determining whether a peptide dendrimer assumes a open (bowl-like) or closed (sphere-like) conformation?

Our results seem to exclude the hypothesis of being exclusively the generation number (for dendrimers with the same number of spacer units), because C1 and B1 present the same generation (and even the same number of residues). Perhaps for higher generations B1, due to *steric hindrance*, might become closed. And it might even exist, as some authors state, a *limit generation* [Maiti 2004; Ballauff 2004], which would be a generation number, specific of every dendrimer at which it is forced to become compact. Still our simulations do not allow to address this question properly. What they do conclude is that for our peptide dendrimers, it is not only the generation number that conditions the conformation adopted.

Another plausible hypothesis concerns the type of branching units used. As the B1-family dendrimers exhibit a greater flexibility than C1, a relation between the use of lysines as branching residues (thus conferring a greater asymmetry to the bifurcation bonds and allowing them to more easily rotate) is not misplaced.

Additionally, the composition of the spacer residues might also condition the conformations, although changing a single core residue in B1 (forming B1H) or

[49] The fact that during this thesis we have referred to B1, B1H, B1HH and B1HHH as the B1-family, and that we are now discussing the classification of peptide dendrimers into families is a mere coincidence. Even before starting the simulations and verify that they share common conformational preferences (which obviously we did not now in advance) we had already assigned them this name.

replacing Amb by histidines (forming B1HH) does not seem to significantly alter the conformations assumed.

The most probable hypothesis (at least in principle) is a conjunction of these (and maybe other) factors without the preponderance of neither of them.

Still, the fact is that after this study, we cannot draw any assertive conclusions on the factors determining the shape adopted by different peptide dendrimers (nor was it our objective), only speculate on it through supported hypothesis. Consequently, we end up with more questions than answers, as often happens in such exploratory work, and demonstrating the potential inherent to this field of research.

In this study, we have also contributed to the discussion around the concept of *dendrimer universal features*. By investigating and comparing the preferential configurations of B1 (or any other B1-family dendrimer) and C1, we are in fact studying the conformations of two "types" of dendrimers with the same global architecture (number of generations and schematic disposition of the residues), but despite that, they present distinct conformational behaviors. Therefore, our studies suggest that dendrimers with the same number of generations do not feature conformational preferences as a direct consequence of dendritic architecture (*universal features*).

Another important feature of the dendrimers studied here, and that might be generalized for all peptide dendrimers (although further investigations are necessary), is that peptide dendrimers seem to be very robust molecules. We base this statement on the fact that, small fluctuations in the amino acid residue composition present little repercussion in terms of the accessible conformational states; this is specially emphasized by the conformational analysis of B1, B1H, B1HH and B1HHH.

Moreover, this implies that despite being impossible to predict, *a priori*, the structure of two peptide dendrimers with markedly different amino acid residue composition, it might be conceivable to infer the preferential structure of peptide dendrimers with similar residue compositions. This can constitute a significant and tangible advantage of peptide dendrimers over proteins in terms of future biological applications.

Furthermore, we have confirmed the suitability of the GROMOS96 53A6 force field to simulate peptide dendrimers, by comparison of our results with the meager available experimental results.

This work demonstrates once more that computer modeling approaches can have a fundamental impact in scientific research, providing the adequate theoretical framework for analyzing experimental results.

As a concluding remark, we must emphasize that the folding of peptide dendrimers was successfully investigated using synergistic approaches that combined different "angles" and methodologies. The results obtained and methodologies tested will greatly facilitate future investigations, and unraveled potential courses of action. The interest that dendrimers continues to arise within the scientific community is a clear indication that "we have not seen the last" of these molecules, and probably not even the "best". Due to their potential to conjugate different areas of science, as peptides or carbohydrates, and their great conformational flexibility (once more evidenced by this study), dendrimers represent a promising prospect for novel knowledge-base applications, alongside with interesting theoretical and conceptual problems.

4.2. Future Perspectives

Based on the high potential exhibited by dendritic molecules in general, and the conformational flexibility that seems to generally characterize peptide dendrimers, further computational and experimental endeavors are to be expected in this field. Specially for molecules combining dendritic architectures with other functional groups. Therefore, we would like to end this thesis with a glance of "what's to come", concerning molecular simulation and other computational approaches of peptide dendrimers.

From our results it is clear that we still cannot fully understand what are the factors conditioning the conformations that dendrimers can adopt in solution; however, we now possess some guidelines to address this question properly.

4. Concluding Remarks

By simulating these same systems, specially B1 and C1, while reducing (B1) or increasing (C1) the size of specific residues, such as the branching units (which in practice equals to iteratively decrease/increase the size of Lyr/Dap), we can determine if the symmetry, or for that effect, the overall size, of the branching residues is what determines the favored shapes assumed by these structures. Also of great help to understand the effective potentialities of peptide dendrimers, is the comprehension of the effects caused by the increase or decrease in the generation number, both for *sphere-like* as well as for *bowl-like* structures. Simulating the fluctuations in the atomic structure as a function of the generation number is, therefore, a key waypoint to shed light on these questions.

It is also important to investigate, with atomic detail, other peptide dendrimers, whose experimental evidences point towards behaviors different from the ones present here. Additionally, the study of peptide dendrimer structures whose amino acid residue compositions have the potential to exhibit secondary structure elements similar to the ones verified in most proteins, are also worth investigating.

Furthermore, to accurately establish a model for the coordination of these dendrimers with aquocobalamin, and establish a structure-function relationship, further studies are needed. Either by simulating, using MM/MD, systems composed of both the aquocobalamin and each peptide dendrimer, or optionally by performing docking studies.

Concerning the more methodological aspects, some computational improvements and implementations might constitute significant gains. For instance, the development of a clustering program specifically prepared to deal with the calculation of RMSD between peptide dendrimers (“permutation problem”).

Appendix A

Chemical structures of different peptide dendrimers

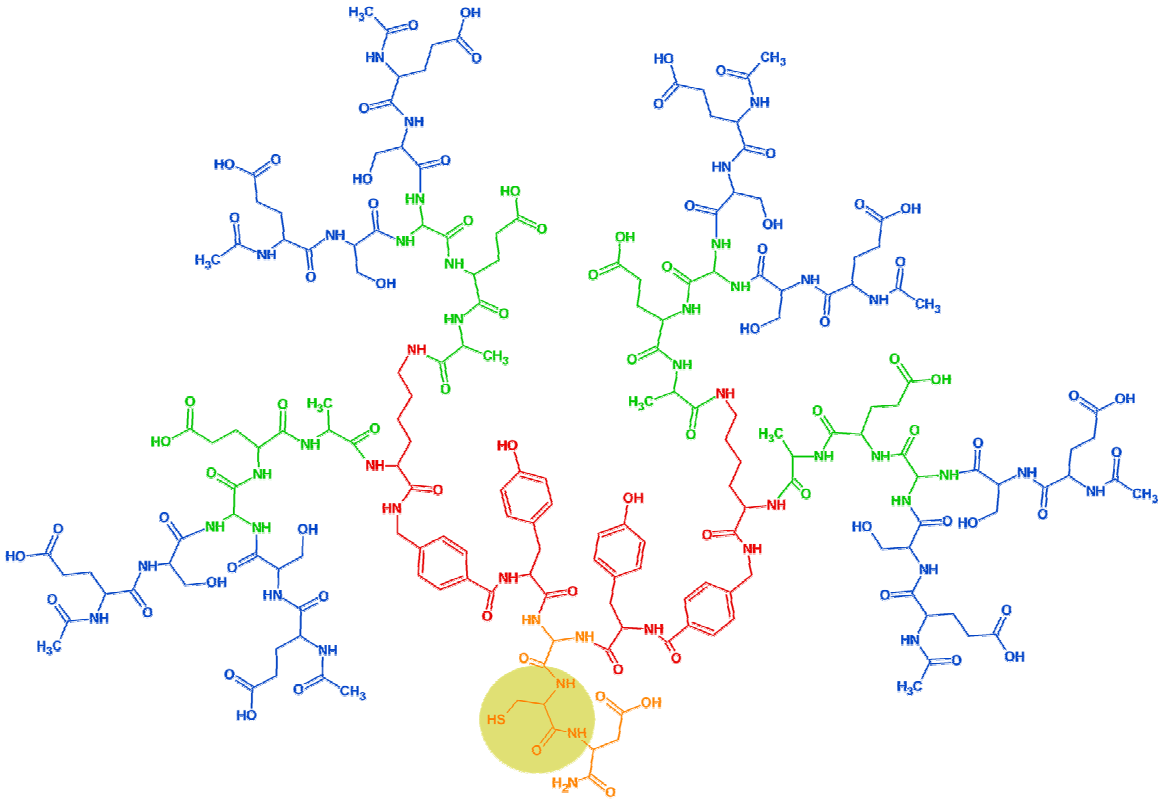


Figure A. 1 – Chemical structure of the B1 dendrimer

Schematic representation of a third generation peptide dendrimer used in the current work. Different generations are highlighted trough the use of a specific color code: G0 – Orange; G1 – Red; G2 – Green; G3 – Blue. The residues presumed to be involved in metal coordination are highlighted by yellow circles. The residue sequence represented corresponds to: $(AceGluSer)_8(DapGluAla)_4(LysAmbTyr)_2DapCysAspNH2$.
Dap=s-2,3-diaminopropanoic acid; **Amb**=4-(aminomethyl)benzoic acid; **Ace**=Acetyl cap; **NH2**=Amine cap; The three letter code representing each proteinogenic amino acid corresponds to the standard nomenclature, but the corresponding names and one-letter codes can still be found in the Abbreviations table presented at the beginning of this thesis.

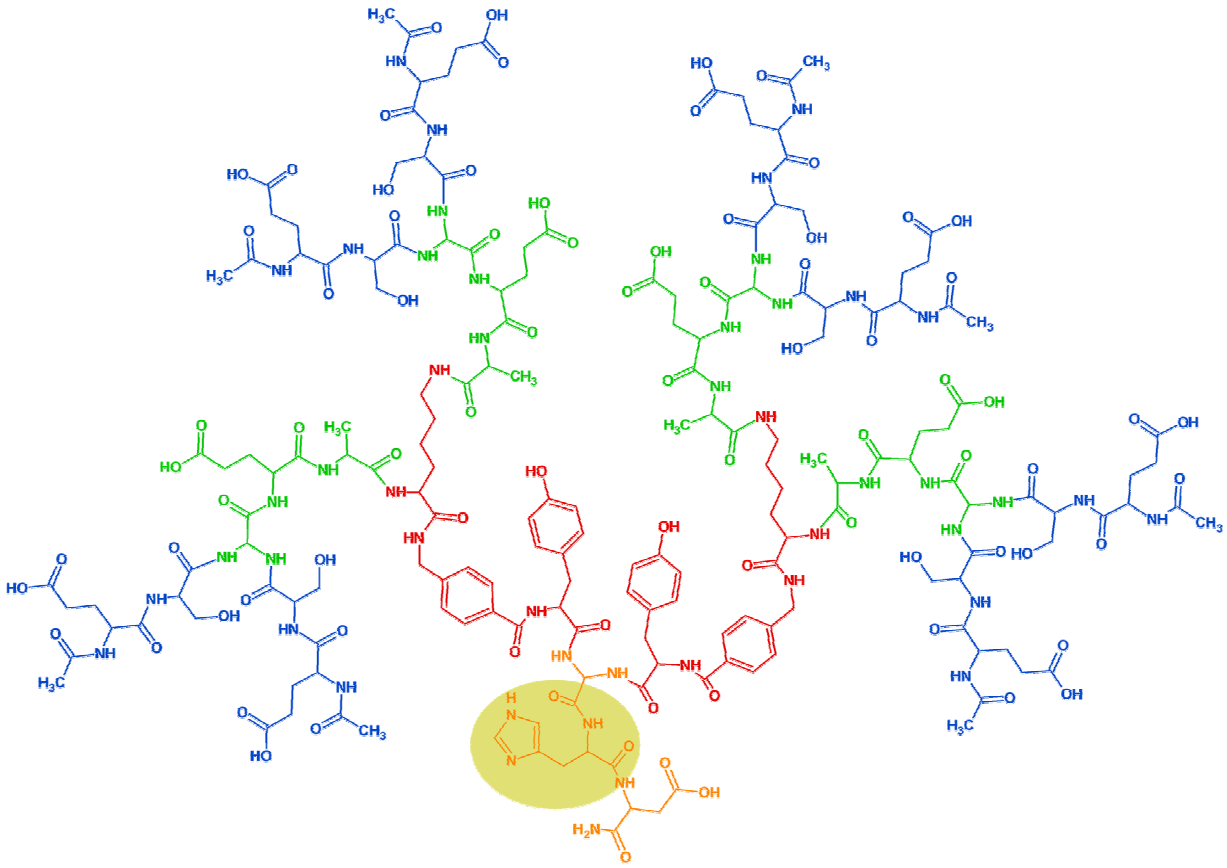


Figure A. 2. – Chemical structure of the B1H dendrimer

Schematic representation of a third generation peptide dendrimer used in the current work. Different generations are highlighted trough the use of a specific color code: G0 – Orange; G1 – Red; G2 – Green; G3 – Blue. The residues presumed to be involved in metal coordination are highlighted by yellow circles. The residue sequence represented corresponds to: $(AceGluSer)_8(DapGluAla)_4(LysAmbTyr)_2DapHisAspNH2$.

Dap=s-2,3-diaminopropanoic acid; **Amb**=4-(aminomethyl)benzoic acid; **Ace**=Acetyl cap; **NH2**=Amine cap; The three letter code representing each proteinogenic amino acid corresponds to the standard nomenclature, but the corresponding names and one-letter codes can still be found in the Abbreviations table presented at the beginning of this thesis.

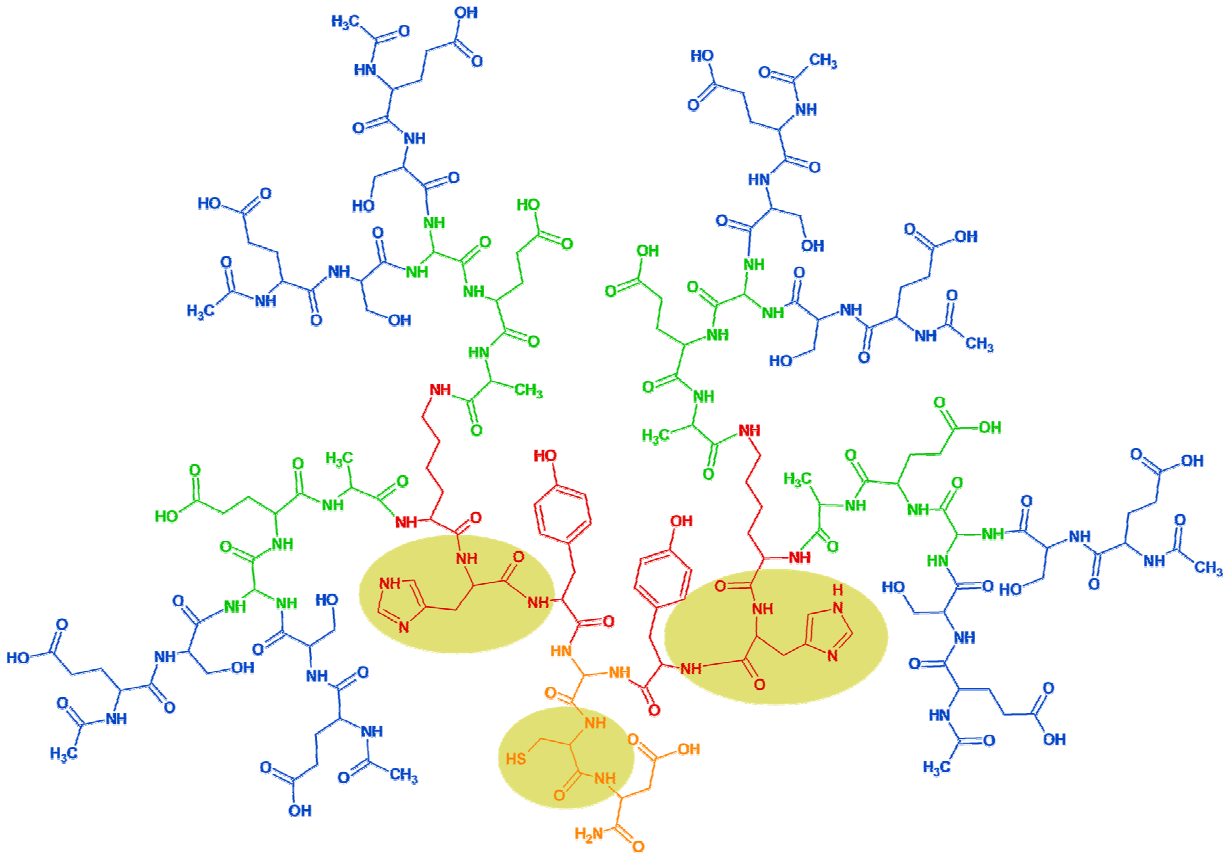


Figure A. 3. – Chemical structure of the B1HH dendrimer

Schematic representation of a third generation peptide dendrimer used in the current work. Different generations are highlighted trough the use of a specific color code: G0 – Orange; G1 – Red; G2 – Green; G3 – Blue. The residues presumed to be involved in metal coordination are highlighted by yellow circles. The residue sequence represented corresponds to: $(AceGluSer)_8(DapGluAla)_4(LysHisTyr)_2DapCysAspNH2$.

Dap=s-2,3-diaminopropanoic acid; **Amb**=4-(aminomethyl)benzoic acid; **Ace**=Acetyl cap; **NH2**=Amine cap; The three letter code representing each proteinogenic amino acid corresponds to the standard nomenclature, but the corresponding names and one-letter codes can still be found in the Abbreviations table presented at the beginning of this thesis.

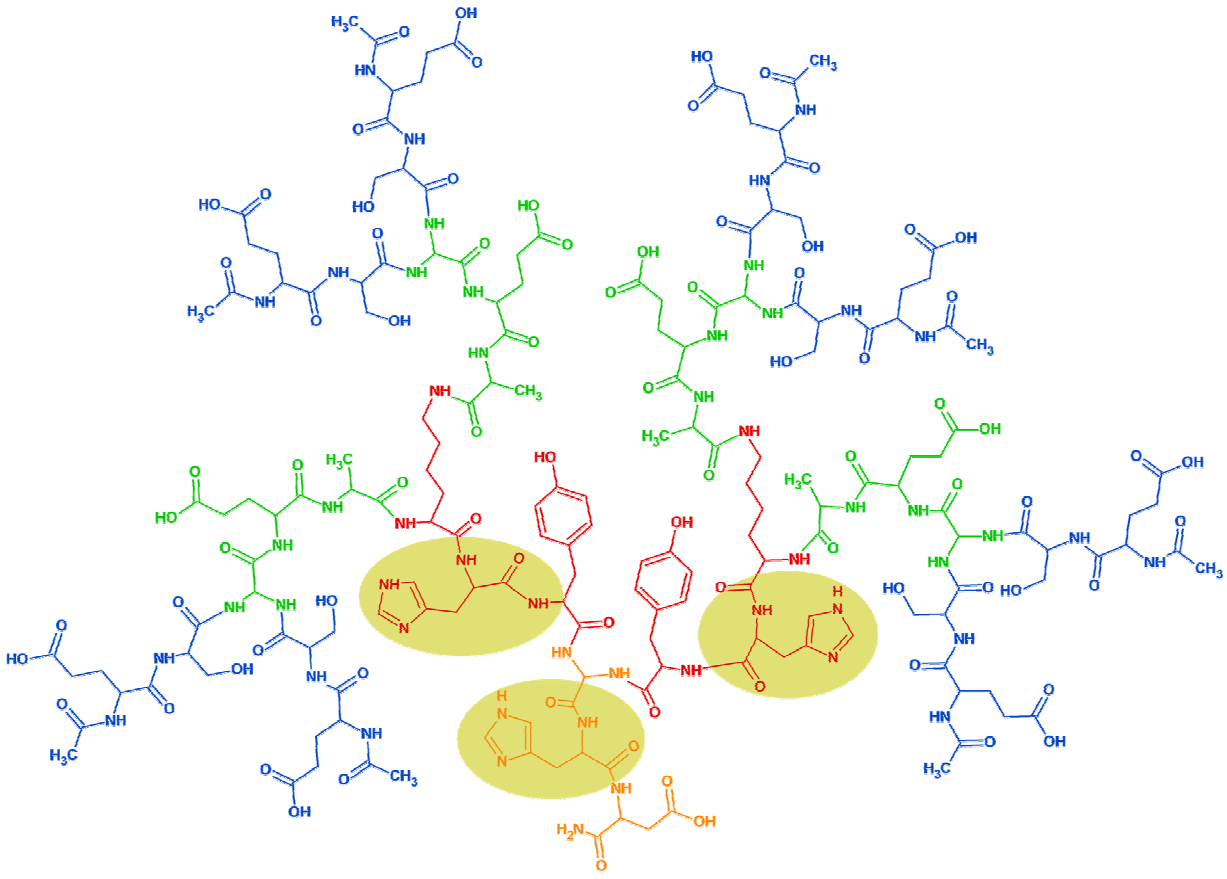


Figure A. 4. – Chemical structure of the B1HHH dendrimer

Schematic representation of a third generation peptide dendrimer used in the current work. Different generations are highlighted trough the use of a specific color code: G0 – Orange; G1 – Red; G2 – Green; G3 – Blue. The residues presumed to be involved in metal coordination are highlighted by yellow circles. The residue sequence represented corresponds to: $(AceGluSer)_8(DapGluAla)_4(LysHisTyr)_2DapHisAspNH2$.

Dap=s-2,3-diaminopropanoic acid; **Amb**=4-(aminomethyl)benzoic acid; **Ace**=Acetyl cap; **NH2**=Amine cap; The three letter code representing each proteinogenic amino acid corresponds to the standard nomenclature, but the corresponding names and one-letter codes can still be found in the Abbreviations table presented at the beginning of this thesis.

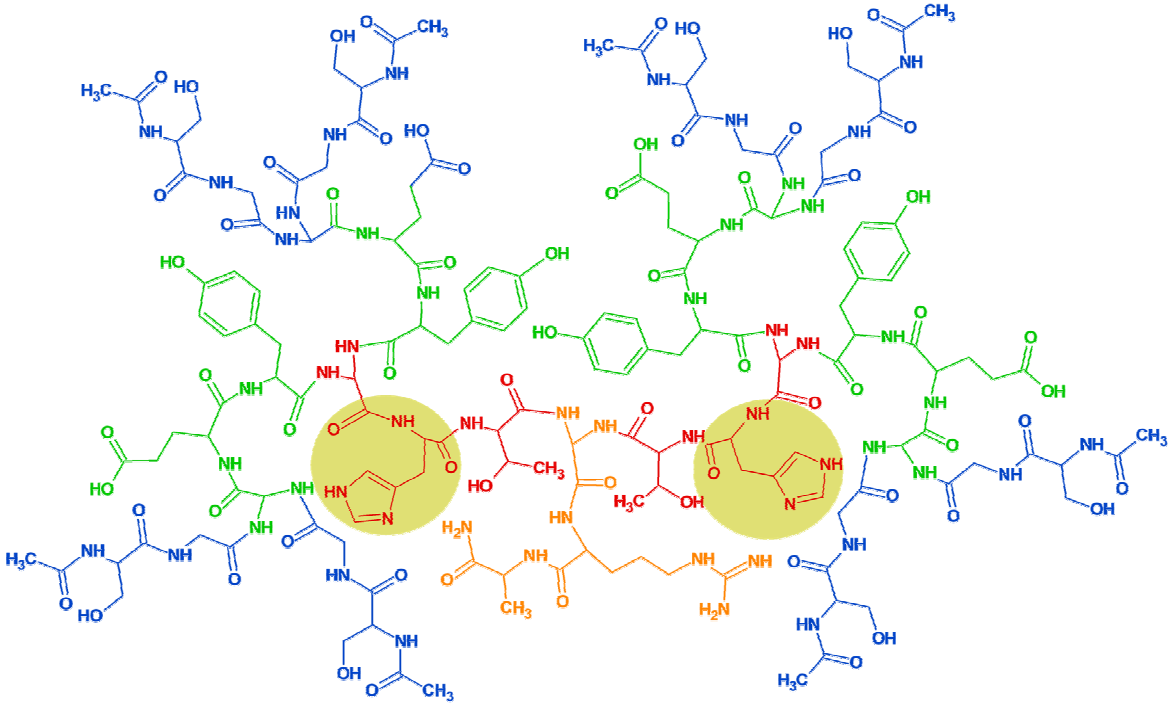


Figure A. 5. – Chemical structure of the C1 dendrimer

Schematic representation of a third generation peptide dendrimer used in the current work. Different generations are highlighted trough the use of a specific color code: G0 – Orange; G1 – Red; G2 – Green; G3 – Blue. The residues presumed to be involved in metal coordination are highlighted by yellow circles. The residue sequence represented corresponds to: $(AceSerGly)_8(DapGluTyr)_4(DapHisThr)_2DapArgAlaNH2$.

Dap=s-2,3-diaminopropanoic acid; **Ace**=Acetyl cap; **NH2**=Amine cap; The three letter code representing each proteinogenic amino acid corresponds to the standard nomenclature, but the corresponding names and one-letter codes can still be found in the Abbreviations table presented at the beginning of this thesis.

Appendix B

GROMOS96 53A6 FF Fragment Parameters

Table B.1 – GROMOS96 53A6 FF Fragment Parameters

Parameter macros employed in the fragments parameterization for the GROMOS96 53A6 Force Field. The values can be found in [Oostenbrink 2004]. The scheme presented must be interpreted having in mind the GROMACS package configuration. The abbreviations gb_x, ga_x, gd_x and gi_x, mean gromos bond (gb), angle (ga), dihedral (gd) and improper (gi); an x accounts for a number that is associated with a certain parameter value in the force field. Therefore, gb_1, is a bond of type 1. The fragments/residues represented are: Nh2 (amino cap), Amb (spacer unit), Dap (branching unit) and Lyr (branching lysine). The order in which the parameters are depicted is as follows: first the values for defining the building blocks that compose each residue (residue database); second the number and placement of the hydrogen atoms (hydrogen database); and last the values characterizing the non-peptidic bonds (special bonds file) between the residue and the other residues it might interact with, which are only apllicable for branching units – Dap and Lyr.

NH2		
Residue Database	**Atoms**	N NT -0.83 0 H1 H 0.415 0 H2 H 0.415 0
	Bonds	N H1 gb_2 N H2 gb_2 -C N gb_9
	Angles	-O -C N ga_33 -CA -C N ga_19 -C N H1 ga_23 -C N H2 ga_23 H1 N H2 ga_24
	Dihedrals	-CA -C N H1 gd_14
	Impropers	-C -O N -CA gi_1 N H1 H2 -C gi_1
Hydrogen Database	2 3 N -C -CA	

Appendix B

DAP		
Residue Database	**Atoms**	N N -0.31000 0 H H 0.31000 0 CA CH1 0.00000 1 CB CH2 0.00000 1 NG N -0.31000 2 HG H 0.31000 2 C C 0.450 3 O O -0.450 3
Residue Database	**Bonds**	N H gb_2 N CA gb_21 CA CB gb_27 CA C gb_27 CB NG gb_21 NG HG gb_2 C O gb_5 C +N gb_10
	Angles	-C N H ga_32 -C N CA ga_31 H N CA ga_18 N CA CB ga_13 N CA C ga_13 CB CA C ga_13 CA CB NG ga_15 CB NG HG ga_11 CA C O ga_30 CA C +N ga_19 O C +N ga_33
	Dihedrals	-CA -C N CA gd_14 -C N CA C gd_39 N CA CB NG gd_34 CA CB NG HG gd_29 N CA C +N gd_40
	Impropers	N -C CA H gi_1 CA N C CB gi_2
Hydrogen Database	1 1 N -C CA 1 1 NG CB CA	
Special Bonds	DAP NG1 GLY C 1 0.140 DAP GLY DAP NG1 THR C 1 0.140 DAP THR DAP NG1 TYR C 1 0.140 DAP TYR DAP NG1 ALA C 1 0.140 DAP ALA DAP NG1 VAL C 1 0.140 DAP VAL DAP NG1 LEU C 1 0.140 DAP LEU DAP NG1 ILEC 1 0.140 DAP ILE DAP NG1 PRO C 1 0.140 DAP PRO DAP NG1 TRP C 1 0.140 DAP TRP DAP NG1 PHE C 1 0.140 DAP PHE DAP NG1 MET C 1 0.140 DAP MET DAP NG1 SER C 1 0.140 DAP SER DAP NG1 CYS C 1 0.140 DAP CYS DAP NG1 GLN C 1 0.140 DAP GLN DAP NG1 ASN C 1 0.140 DAP ASN DAP NG1 LYS C 1 0.140 DAP LYS DAP NG1 ARG C 1 0.140 DAP ARG DAP NG1 HIS C 1 0.140 DAP HIS DAP NG1 ASP C 1 0.140 DAP ASP DAP NG1 GLU C 1 0.140 DAP GLU DAP NG1 AMB C 1 0.140 DAP AMB	

AMB		
Residue Database	Atoms	N N -0.31000 0 H H 0.31000 0 CA CH2 0.00000 1 CB C 0.00000 1 CG1 C -0.14000 2 HG1 HC 0.14000 2 CG2 C -0.14000 3 HG2 HC 0.14000 3 CD1 C -0.14000 4 HD1 HC 0.14000 4 CD2 C -0.14000 5 HD2 HC 0.14000 5 CE C 0.00000 6 C C 0.450 7 O O -0.450 7
Residue Database	Bonds	N H gb_2 N CA gb_21 CA CB gb_27 CB CG1 gb_16 CB CG2 gb_16 CG1 HG1 gb_3 CG1 CD1 gb_16 CG2 HG2 gb_3 CG2 CD2 gb_16 CD1 HD1 gb_3 CD1 CE gb_16 CD2 HD2 gb_3 CD2 CE gb_16 CE C gb_27 C O gb_5 C +N gb_10
	Exclusions	CA HG1 CA HG2 CA CD1 CA CD2 CB HD1 CB HD2 CB CE CG1 HG2 CG1 CD2 CG1 C HG1 CG2 HG1 HD1 HG1 CE CG2 CD1 CG2 C HG2 HD2 HG2 CE CD1 HD2 HD1 CD2 HD1 C HD2 C
	Angles	-C N H ga_32 -C N CA ga_31 H N CA ga_18 N CA CB ga_13 CA CB CG1 ga_27 CA CB CG2 ga_27 CG1 CB CG2 ga_27 CB CG1 HG1 ga_25 CB CG1 CD1 ga_27 HG1 CG1 CD1 ga_25

Continues on next page

Residue Database	**Angles**	CB CG2 HG2 ga_25 CB CG2 CD2 ga_27 HG2 CG2 CD2 ga_25 CG1 CD1 HD1 ga_25 CG1 CD1 CE ga_27 HD1 CD1 CE ga_25 CG2 CD2 HD2 ga_25 CG2 CD2 CE ga_27 HD2 CD2 CE ga_25 CD1 CE CD2 ga_27 CD1 CE C ga_27 CD2 CE C ga_27 CE C O ga_30 CE C +N ga_19 O C +N ga_33
	Dihedrals	-CA -C N CA gd_14 CD1 CE C O gd_10 N CA CB CG1 gd_40
	Impropers	N -C CA H gi_1 CB CG1 CG2 CA gi_1 CB CG1 CD1 CE gi_1 CB CG2 CD2 CE gi_1 CG1 CB CG2 CD2 gi_1 CG1 CB CD1 HG1 gi_1 CG1 CD1 CE CD2 gi_1 CG2 CB CG1 CD1 gi_1 CG2 CB CD2 HG2 gi_1 CG2 CD2 CE CD1 gi_1 HD1 CG1 CE CD1 gi_1 HD2 CG2 CE CD2 gi_1 CE CD1 CD2 C gi_1 C CE +N O gi_1
Hydrogen Database		1 1 N -C CA 1 1 CG1 CB CD1 1 1 CG2 CB CD2 1 1 CD1 CG1 CE 1 1 CD2 CG2 CE

LYR		
Residue Database	**Atoms**	N N -0.31000 0 H H 0.31000 0 CA CH1 0.00000 1 CB CH2 0.00000 1 CG CH2 0.00000 2 CD CH2 0.00000 2 CE CH2 0.00000 3 NZ N -0.31000 3 HZ H 0.31000 3 C C 0.450 4 O O -0.450 4
	Bonds	N H gb_2 N CA gb_21 CA CB gb_27 CA C gb_27 CB CG gb_27 CG CD gb_27 CD CE gb_27 CE NZ gb_21

Continues on next page

Residue Database	**Bonds**	NZ HZ gb_2 C O gb_5 C +N gb_10
	Angles	-C N H ga_32 -C N CA ga_31 H N CA ga_18 N CA CB ga_13 N CA C ga_13 CB CA C ga_13 CA CB CG ga_15 CB CG CD ga_15 CG CD CE ga_15 CD CE NZ ga_13 CE NZ HZ ga_18 CA C O ga_30 CA C +N ga_19 O C +N ga_33
	Dihedrals	-CA -C N CA gd_14 -C N CA C gd_39 N CA CB CG gd_34 N CA C +N gd_40 CA CB CG CD gd_34 CB CG CD CE gd_34 CG CD CE NZ gd_34
	Impropers	N -C CA H gi_1 CA N C CB gi_2 C CA +N O gi_1
Hydrogen Database	1 1 N -C CA 1 1 NZ CE CD	
Special Bonds	LYR NZ1 GLY C 1 0.140 LYR GLY LYR NZ1 THR C 1 0.140 LYR THR LYR NZ1 TYR C 1 0.140 LYR TYR LYR NZ1 ALA C 1 0.140 LYR ALA LYR NZ1 VAL C 1 0.140 LYR VAL LYR NZ1 LEU C 1 0.140 LYR LEU LYR NZ1 ILE C 1 0.140 LYR ILE LYR NZ1 PRO C 1 0.140 LYR PRO LYR NZ1 TRP C 1 0.140 LYR TRP LYR NZ1 PHE C 1 0.140 LYR PHE LYR NZ1 MET C 1 0.140 LYR MET LYR NZ1 SER C 1 0.140 LYR SER LYR NZ1 CYS C 1 0.140 LYR CYS LYR NZ1 GLN C 1 0.140 LYR GLN LYR NZ1 ASN C 1 0.140 LYR ASN LYR NZ1 LYS C 1 0.140 LYR LYS LYR NZ1 ARG C 1 0.140 LYR ARG LYR NZ1 HIS C 1 0.140 LYR HIS LYR NZ1 ASP C 1 0.140 LYR ASP LYR NZ1 GLU C 1 0.140 LYR GLU LYR NZ1 AMB C 1 0.140 LYR AMB	

Appendix C

Scatterplots and time series

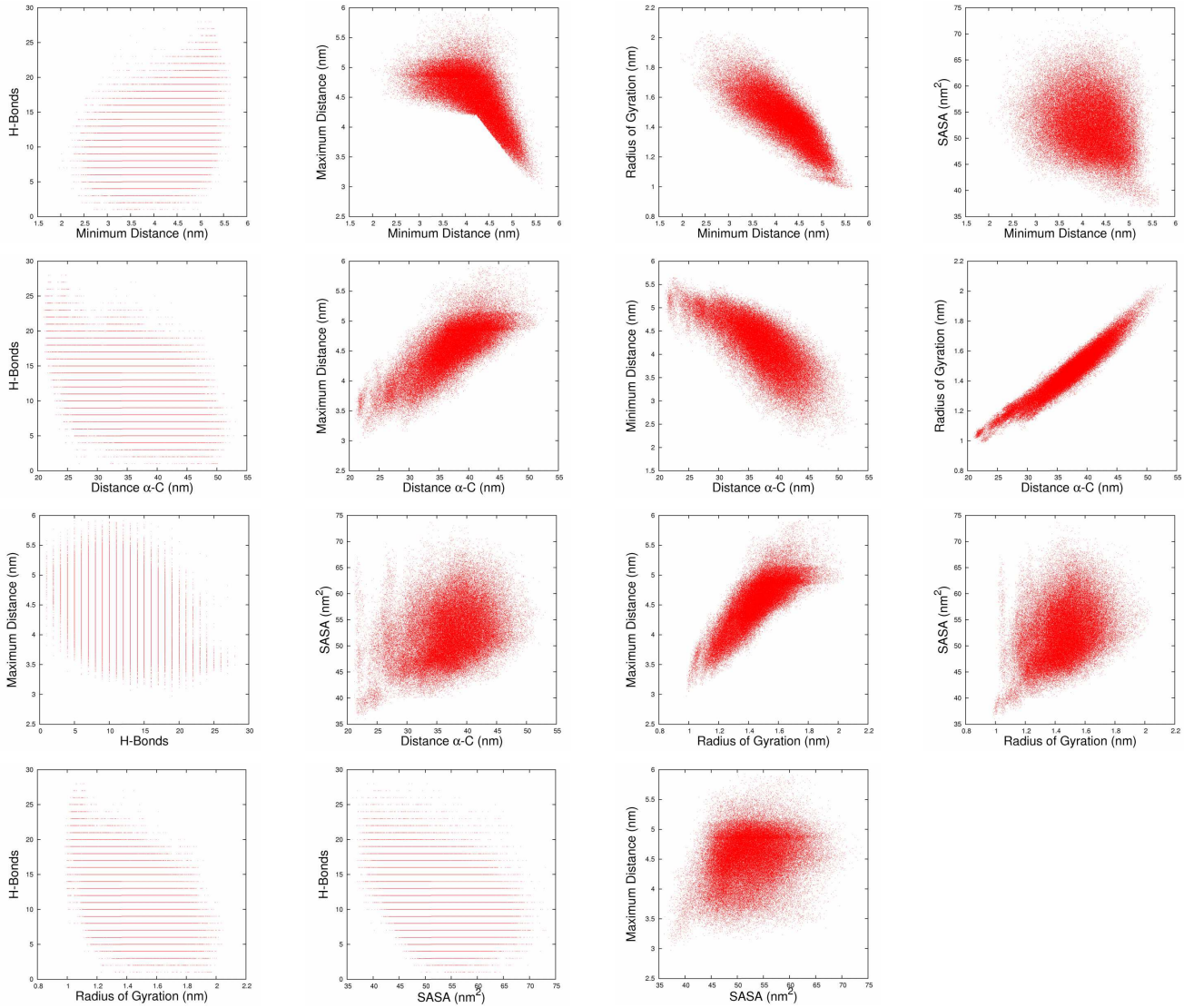


Figure C. 1 – Scatterplots combining multiple properties calculated for B1H.

Scatterplots displaying the relation between pairs of properties computed for the B1H peptide dendrimer. The values were determined using all replicates and the entire 100 ns trajectories. The properties considered are: the total number of hydrogen bonds (#H-Bonds), the solvent accessible surface area (SASA), the radius of gyration (R_g), the sum of the distance among the branching units alpha carbons (Distance α-C), the minimum distance with the periodic image (Minimum Distance), and the distance between the two farthest atoms (Maximum Distance).

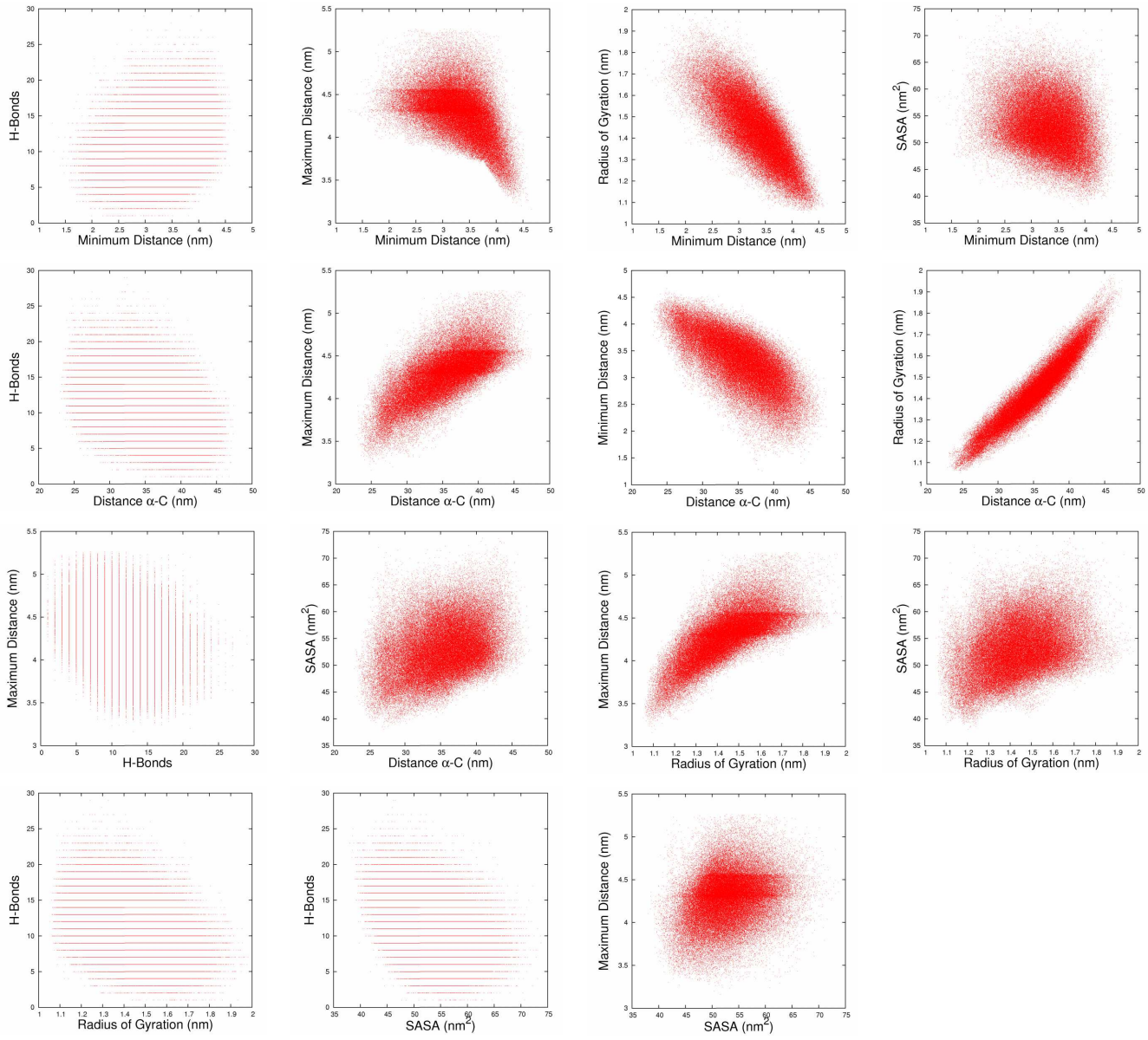


Figure C. 2 – Scatterplots combining multiple properties calculated for B1HH.

Scatterplots displaying the relation between pairs of properties computed for the B1HH peptide dendrimer. The values were determined using all replicates and the entire 100 ns trajectories. The properties considered are: the total number of hydrogen bonds (#H-Bonds), the solvent accessible surface area (SASA), the radius of gyration (R_g), the sum of the distance among the branching units alpha carbons (Distance α-C), the minimum distance with the periodic image (Minimum Distance), and the distance between the two farthest atoms (Maximum Distance).

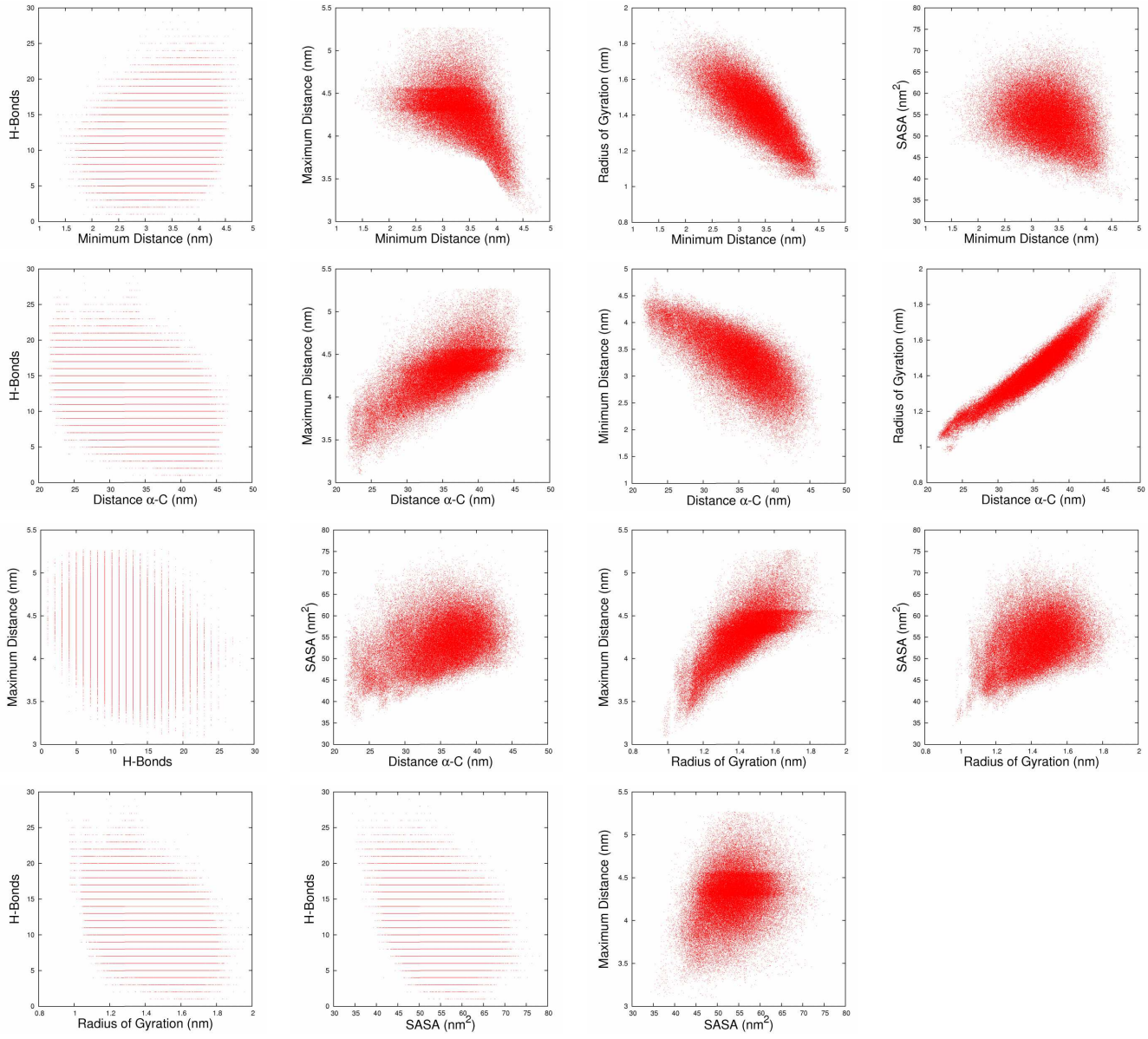


Figure C. 3 – Scatterplots combining multiple properties calculated for B1HHH.

Scatterplots displaying the relation between pairs of properties computed for the B1HHH peptide dendrimer. The values were determined using all replicates and the entire 100 ns trajectories. The properties considered are: the total number of hydrogen bonds (#H-Bonds), the solvent accessible surface area (SASA), the radius of gyration (Rg), the sum of the distance among the branching units alpha carbons (Distance α-C), the minimum distance with the periodic image (Minimum Distance), and the distance between the two farthest atoms (Maximum Distance).

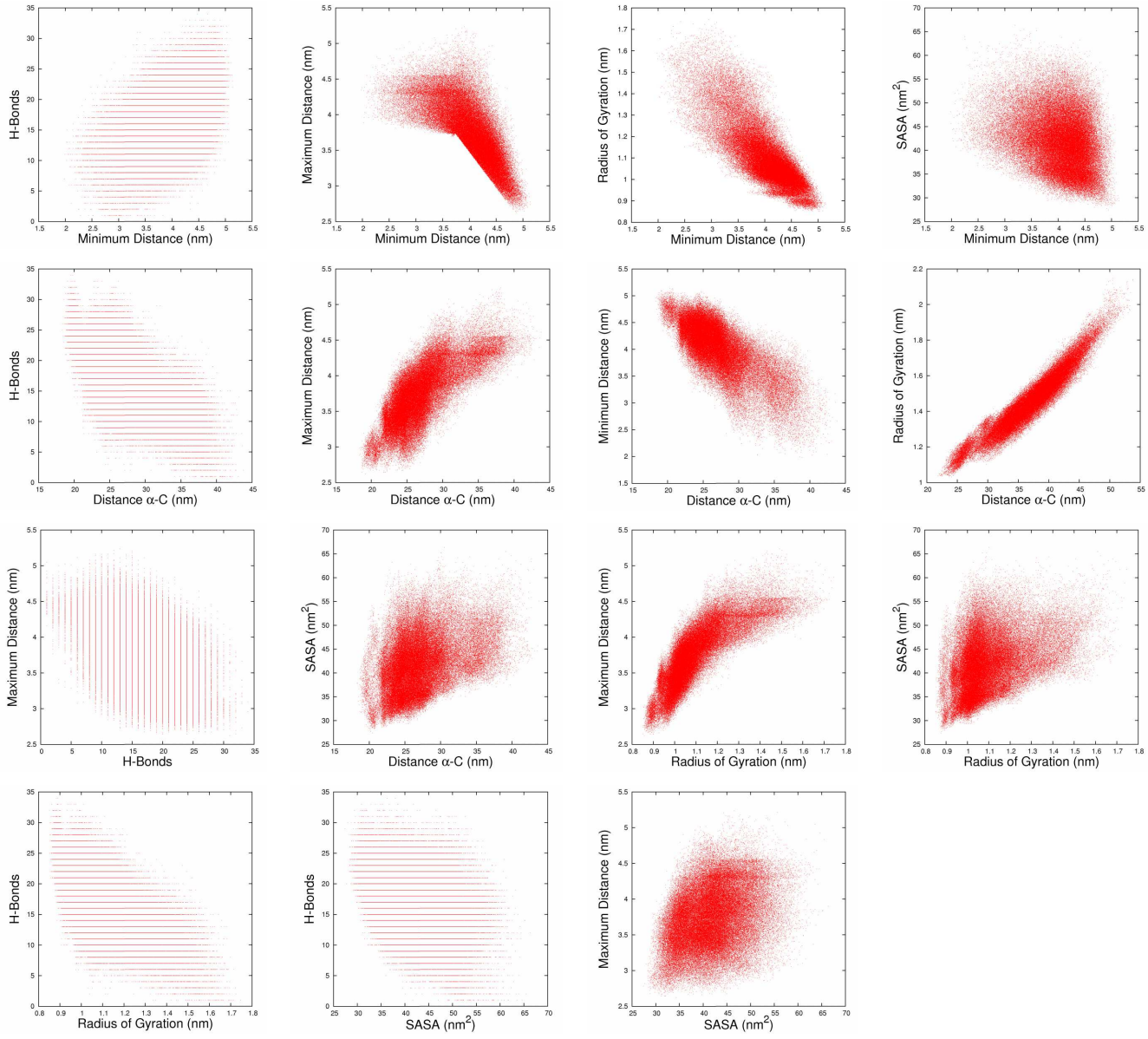


Figure C. 4 – Scatterplots combining multiple properties calculated for C1.

Scatterplots displaying the relation between pairs of properties computed for the C1 peptide dendrimer. The values were determined using all replicates and the entire 100 ns trajectories. The properties considered are: the total number of hydrogen bonds (#H-Bonds), the solvent accessible surface area (SASA), the radius of gyration (R_g), the sum of the distance among the branching units alpha carbons (Distance α-C), the minimum distance with the periodic image (Minimum Distance), and the distance between the two farthest atoms (Maximum Distance).

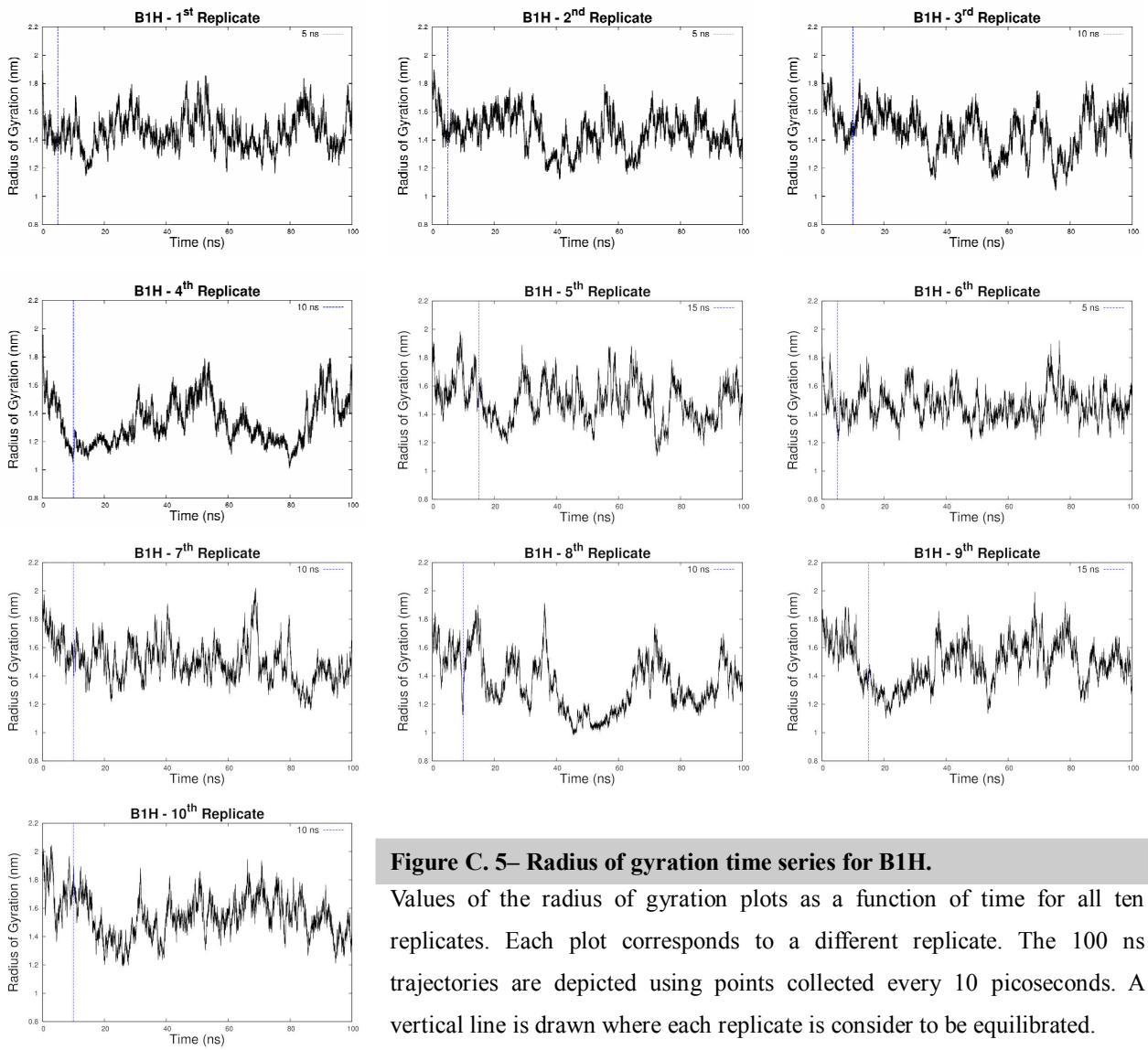


Figure C. 5– Radius of gyration time series for B1H.

Values of the radius of gyration plots as a function of time for all ten replicates. Each plot corresponds to a different replicate. The 100 ns trajectories are depicted using points collected every 10 picoseconds. A vertical line is drawn where each replicate is consider to be equilibrated.

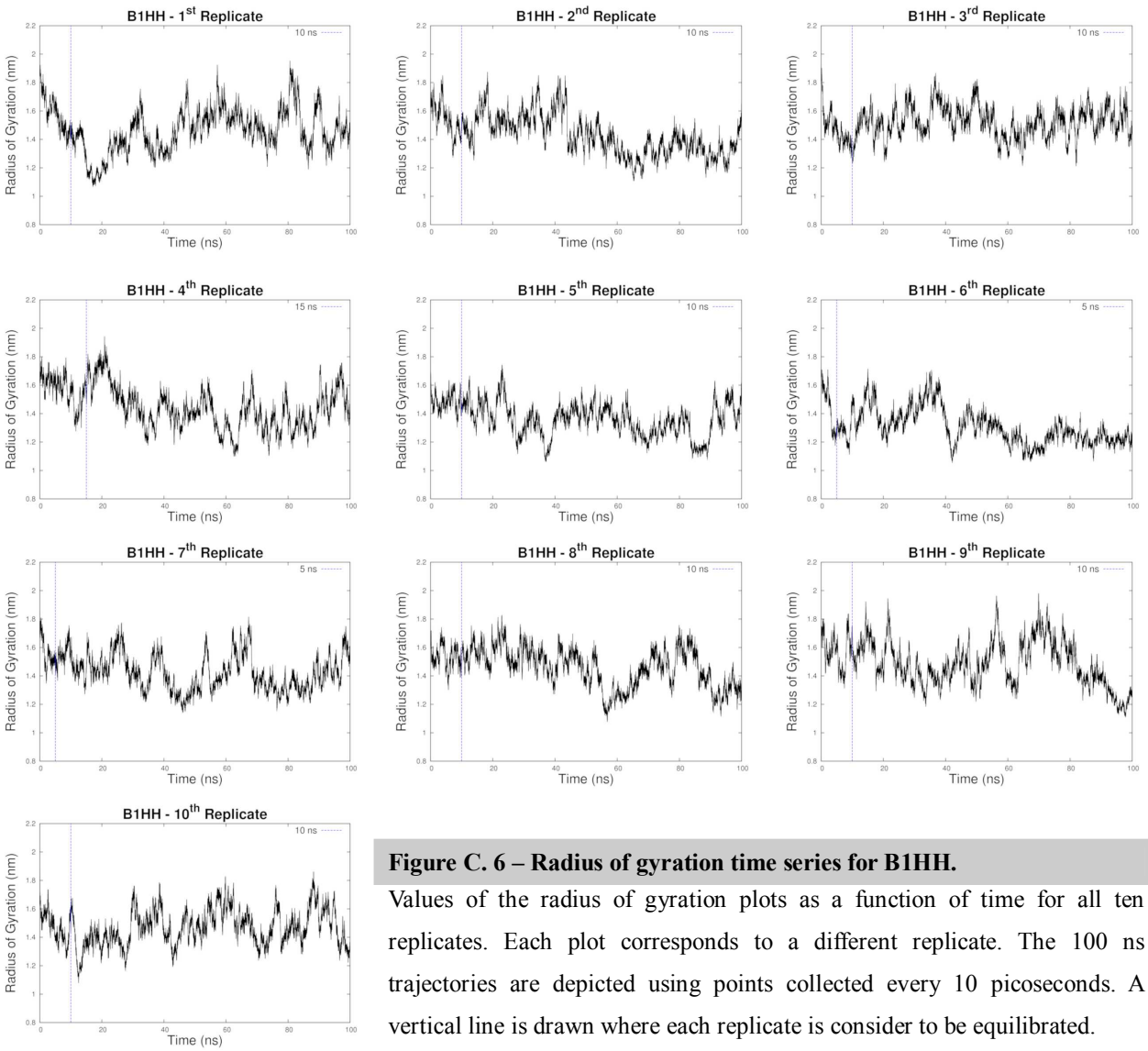


Figure C. 6 – Radius of gyration time series for B1HH.

Values of the radius of gyration plots as a function of time for all ten replicates. Each plot corresponds to a different replicate. The 100 ns trajectories are depicted using points collected every 10 picoseconds. A vertical line is drawn where each replicate is consider to be equilibrated.

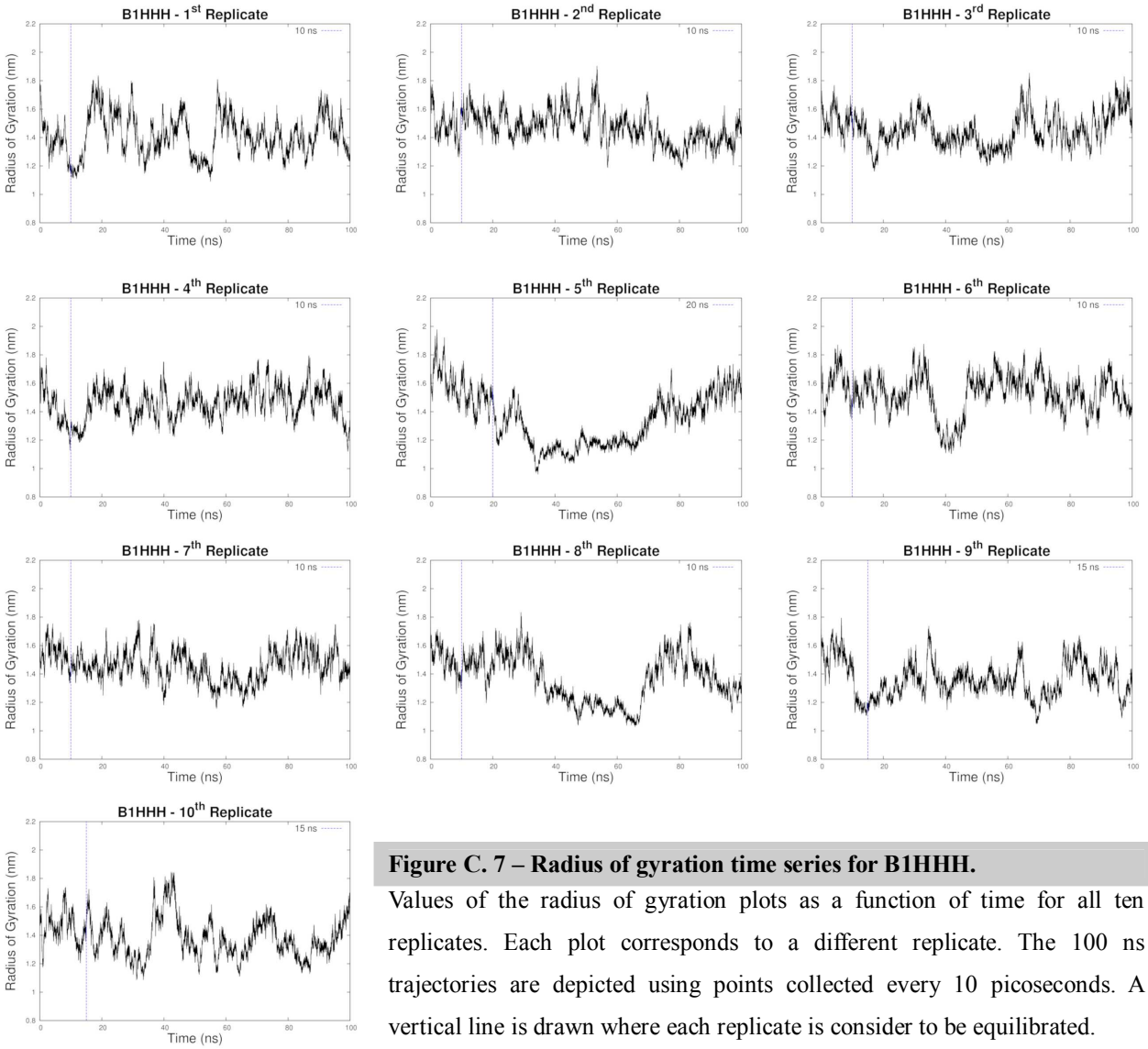


Figure C. 7 – Radius of gyration time series for B1HHH.

Values of the radius of gyration plots as a function of time for all ten replicates. Each plot corresponds to a different replicate. The 100 ns trajectories are depicted using points collected every 10 picoseconds. A vertical line is drawn where each replicate is consider to be equilibrated.

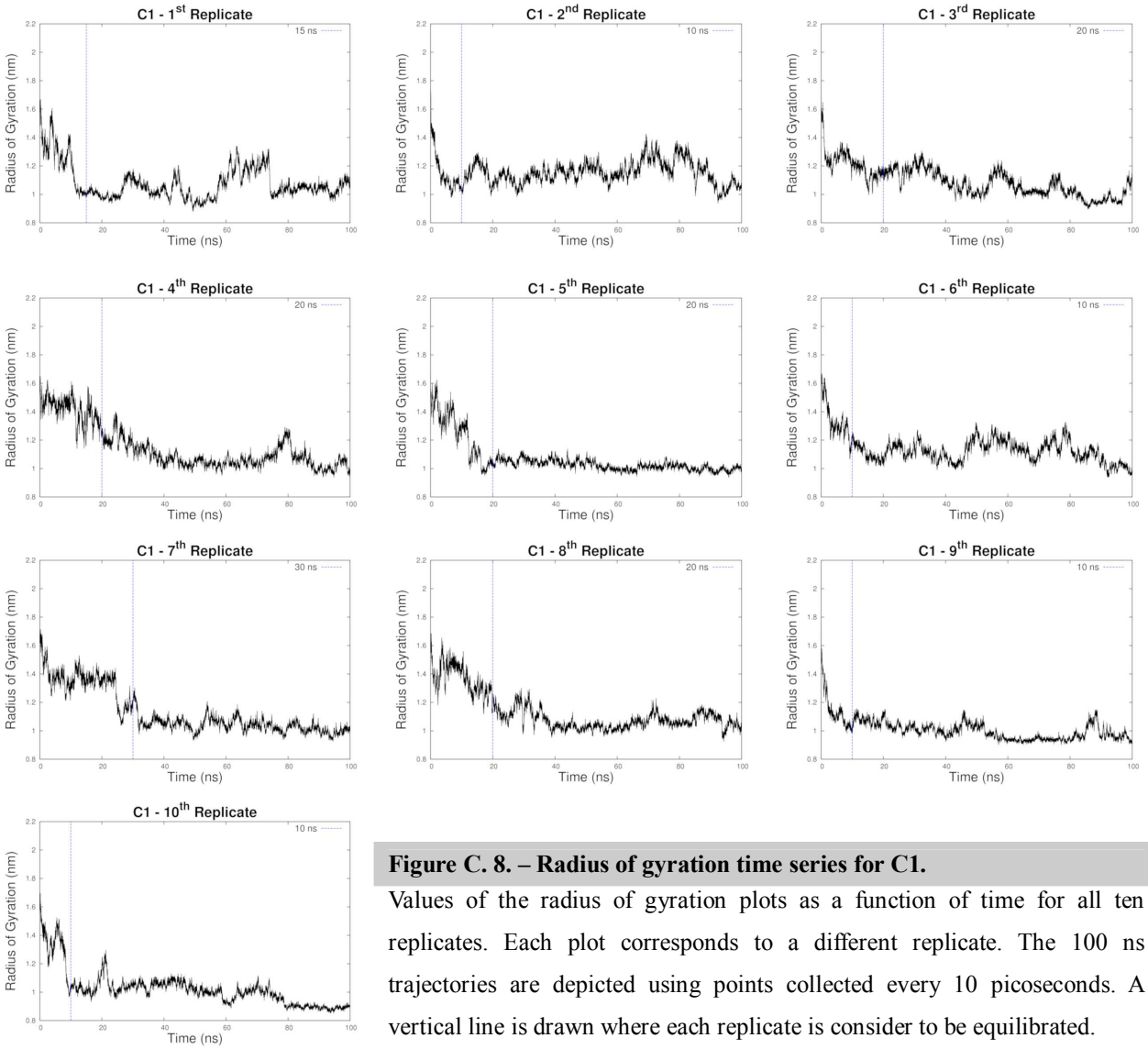


Figure C. 8. – Radius of gyration time series for C1.

Values of the radius of gyration plots as a function of time for all ten replicates. Each plot corresponds to a different replicate. The 100 ns trajectories are depicted using points collected every 10 picoseconds. A vertical line is drawn where each replicate is consider to be equilibrated.

Appendix D

Autocorrelation Functions

Molecular dynamic simulations generate data representing the molecular conformations of a simulated system that are connected in time; therefore, we can use MD to determine time-dependent properties.

Suppose we have two different sets of data values (quantities), A and B, obtained from an MD trajectory at specific times, and we intend to determine what correlation exists between them. Accordingly, the correlation between A and B is measured resorting to statistics [Allen 1989], i.e., through the respective *correlation coefficient*, C_{AB}, defined as,

$$C_{AB} = \frac{\langle \delta A \, \delta B \rangle_{ens}}{\sigma(A)\sigma(B)}, \qquad \text{Equation D.1.}$$

where δA and δB represent the respective deviations from the ensemble averages (depicted as $\langle A \rangle_{ens}$ and $\langle B \rangle_{ens}$), $\delta A = A - \langle A \rangle_{ens}$ and $\delta B = B - \langle B \rangle_{ens}$; and $\sigma(A)$ and $\sigma(B)$ the root mean square deviation for dynamic varible A and B (*standard deviation*), respectivelly. $\sigma(A)$ is thus defined as

$$\sigma(A) = \sqrt{\langle A^2 \rangle_{ens} - \langle A \rangle_{ens}^2} = \sqrt{\langle \delta A^2 \rangle_{ens}}, \qquad \text{Equation D.2.}$$

and the same is true for $\sigma(B)$.

If we estimate A and B at two different times we obtain their correlation function in relation to time, t. A function known as the *time correlation function*, $C_{AB}(t)$.

Moreover, when *A* and *B* are the same quantity the correlation function is normally named *autocorrelation function*, C_{AA} (as opose to *cross-correlation function* when $A \neq B$).

A *time autocorrelation function*, $C_{AA}(t)$, will measure the correlation of property *A(t)* to itself at two different times, separated by the time interval *t*, and averaged over the whole trajectory.

If we consider that the *non-normalized time autocorrelation function*, $C'_{AA}(t)$, is defined as,

$$C'_{AA}(t) = \langle \delta A(t)\, \delta A(0) \rangle_{ens}, \qquad \text{Equation D.3.}$$

expressing the deviation from the essemble average among *A* at time *t* and *A* at time zero; we can write the corresponding time autocorrelation function as,

$$C_{AA}(t) = \frac{C'_{AA}(t)}{\sigma^2(A)} = \frac{C'_{AA}(t)}{\langle A^2 \rangle_{ens} - \langle A \rangle^2_{ens}}. \qquad \text{Equation D.4.}$$

Therefore, we can roughtly consider $C_{AA}(t)$ as the similarity between observations as a function of the time separation between them. In this way, to investigate the existence of periodic events becomes easier. An autocorrelation function has the initial value of 1 and normally converges to 0 at later times. The time taken by a system to lose the "memory", or more properly the correlation, with regard to a certain property, is named *relaxation time* (or sometimes *correlation time*). To determine the correlation function (or any other property) of a certain system under equilibrium conditions, the length of the simulation must be significativelly longer than the relaxation time.

It is important to mention two relevant properties of time autocorrelation functions: they are *reversible*, i.e., $C_A(t) = C_A(-t)$; and they are *stationary*, meaning that, $\langle A(\tau + t) A(\tau) \rangle = \langle A(t) A(0) \rangle$ [Becker 2001].

Appendix D

In pratice, the computation of $C_{AA}(t)$ involves averaging over diffent time intervals in the trajectory, i.e., averaging the terms $\langle A(\tau+t)A(\tau)\rangle$ while sliding τ along the trajectory [Allen 1989; Becker 2001].

For other discussions on concepts regarding correlation functions, the reader is directed to some excellent books by Allen & Tildesley [Allen 1989], Berendsen [Berendsen 2007] and Leach [Leach 2001].

An important aspect of time autocorrelation functions that is of particular importance in the present work, is that the calculation of the average time autocorrelation functions will depend on the property being considered. For instance, to determine the average time autocorrelation function for the radius of gyration (R_g) over all the replicates used, we must resort to Equation D.4., express it taking all replicates into account.

Therefore, we star by operating on the non-normalized time autocorrelation function, representing it as:

$$C'_{AA}(t) = \langle \delta A(t)\, \delta A(0)\rangle_{ens} = \left\langle \left[A(t) - \langle A\rangle_{ens} \right] \left[A(0) - \langle A\rangle_{ens} \right] \right\rangle_{ens} . \qquad \text{Equation D.5.}$$

For the same ensemble we can consider that, $\langle A(t)\rangle_{ens} = \langle A(0)\rangle_{ens}$, and we can even write that,

$$C'_{AA}(t) = \langle A(t)A(0)\rangle_{ens} - \langle A\rangle^2_{ens} . \qquad \text{Equation D.6.}$$

If we define the total number of replicates as R, and r as an indice that comprises all the replicates, $r = 1,2,...,\mathrm{R}$, then we can define the average over all replicates as,

$$\langle A\rangle_{ens} = \frac{1}{\mathrm{R}} \sum_r \langle A\rangle_r ;$$

and the average over the square of the property as

$$\langle A^2\rangle_{ens} = \frac{1}{\mathrm{R}} \sum_r \langle A^2\rangle_r .$$

If we now rewrite the average over a time interval as a function of the number of replicates,

$$\langle A(t)A(0)\rangle_{ens} = \frac{1}{\mathrm{R}}\sum_r \langle A(t)A(0)\rangle_r ,$$

we can define the time autocorrelation function average over all replicates as:

$$C_{AA}(t) = \frac{C'_{AA}(t)}{\sigma^2(A)} = \frac{\langle A(t)A(0)\rangle_{ens} - \langle A\rangle^2_{ens}}{\langle A^2\rangle_{ens} - \langle A\rangle^2_{ens}} = \frac{\frac{1}{\mathrm{R}}\sum_r \langle A(t)A(0)\rangle_r - \left[\frac{1}{\mathrm{R}}\sum_r \langle A\rangle_r\right]^2}{\frac{1}{\mathrm{R}}\sum_r \langle A^2\rangle_r - \left[\frac{1}{\mathrm{R}}\sum_r \langle A\rangle_r\right]^2} .$$

Equation D.7.

Which, despite being somewhat involved, is the adequated form to calculate the average over all replicates for R_g using the computationally acessible data.

Particularizing for the radius of gyration we obtain the equation used in this thesis:

$$C_{\mathrm{R_g}}(t) = \frac{\frac{1}{\mathrm{R}}\sum_r \langle \mathrm{R_g}(t)\mathrm{R_g}(0)\rangle_r - \left[\frac{1}{\mathrm{R}}\sum_r \langle \mathrm{R_g}\rangle_r\right]^2}{\frac{1}{\mathrm{R}}\sum_r \langle \mathrm{R_g}^2\rangle_r - \left[\frac{1}{\mathrm{R}}\sum_r \langle \mathrm{R_g}\rangle_r\right]^2} .$$

Equation D.8.

Appendix E

Radius of gyration time autocorrelation functions

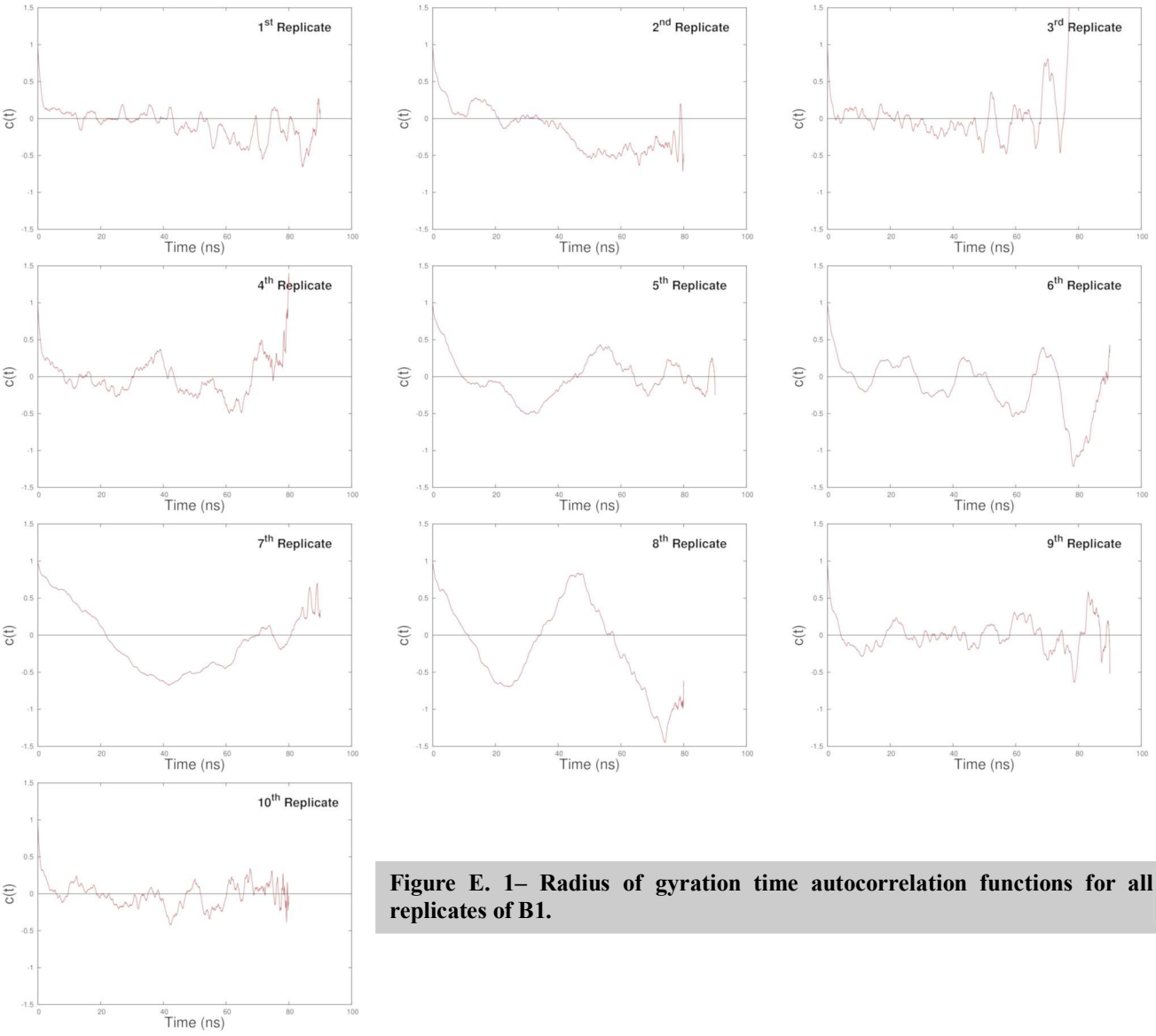


Figure E. 1– Radius of gyration time autocorrelation functions for all replicates of B1.

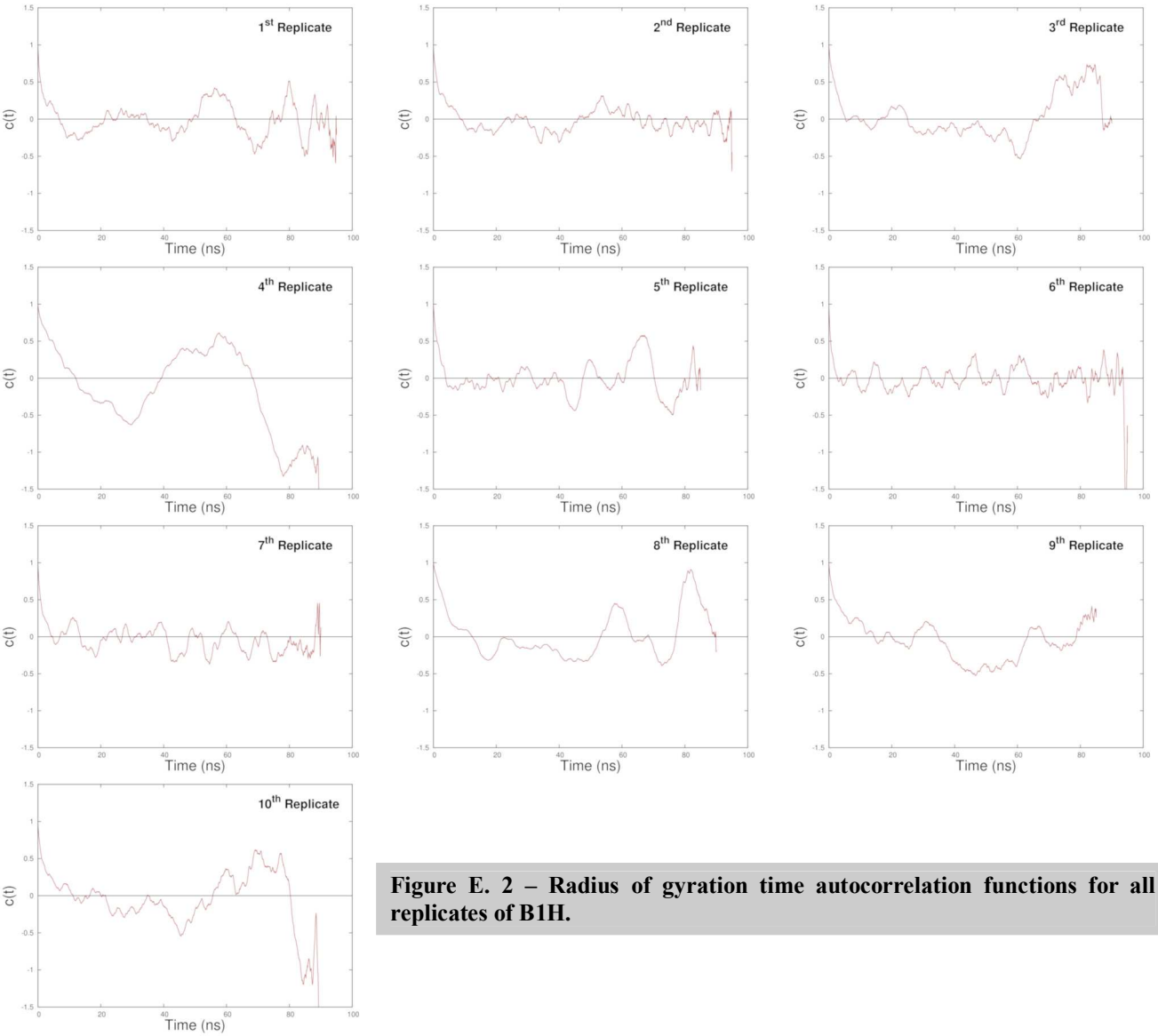


Figure E. 2 – Radius of gyration time autocorrelation functions for all replicates of B1H.

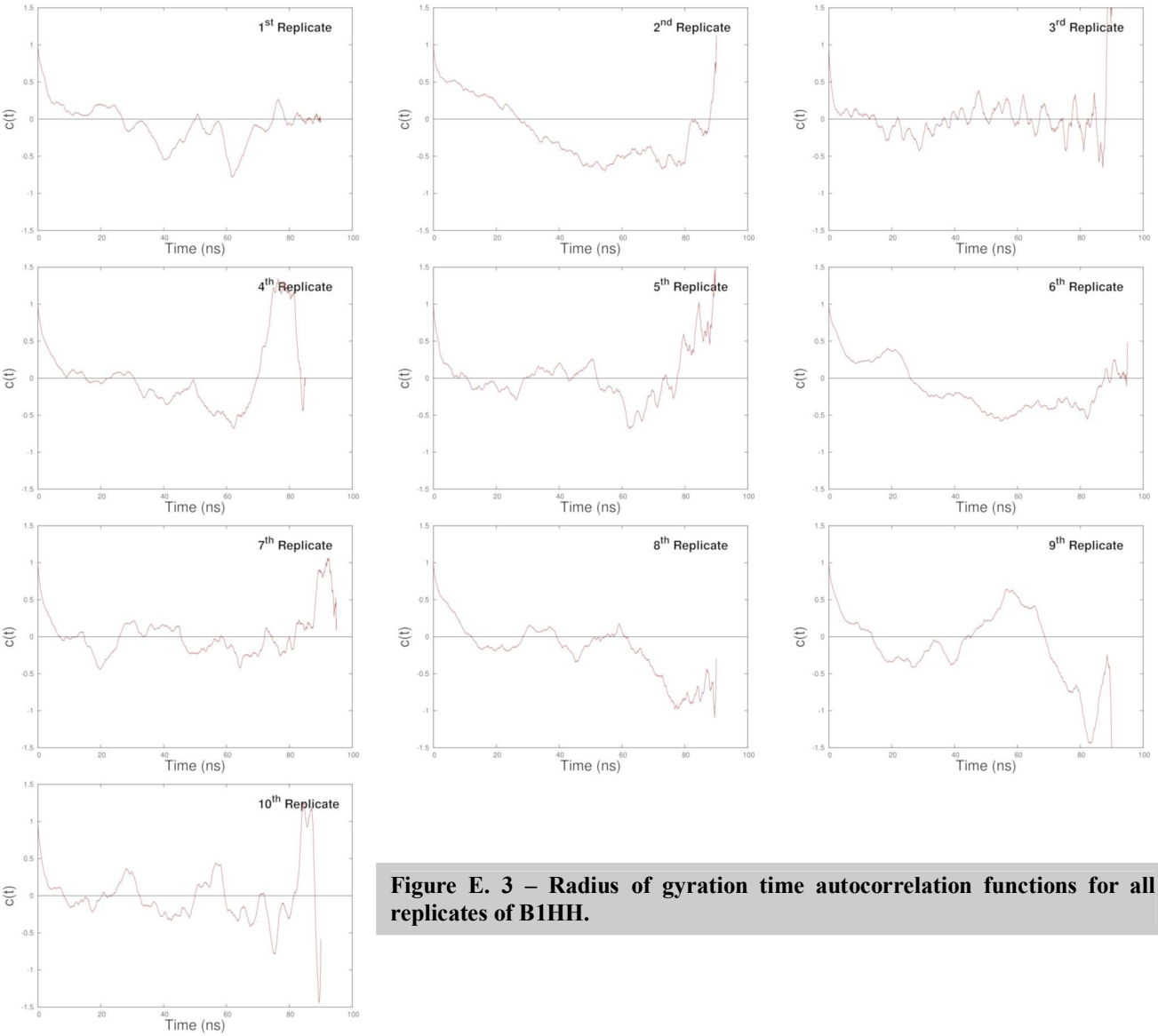


Figure E. 3 – Radius of gyration time autocorrelation functions for all replicates of B1HH.

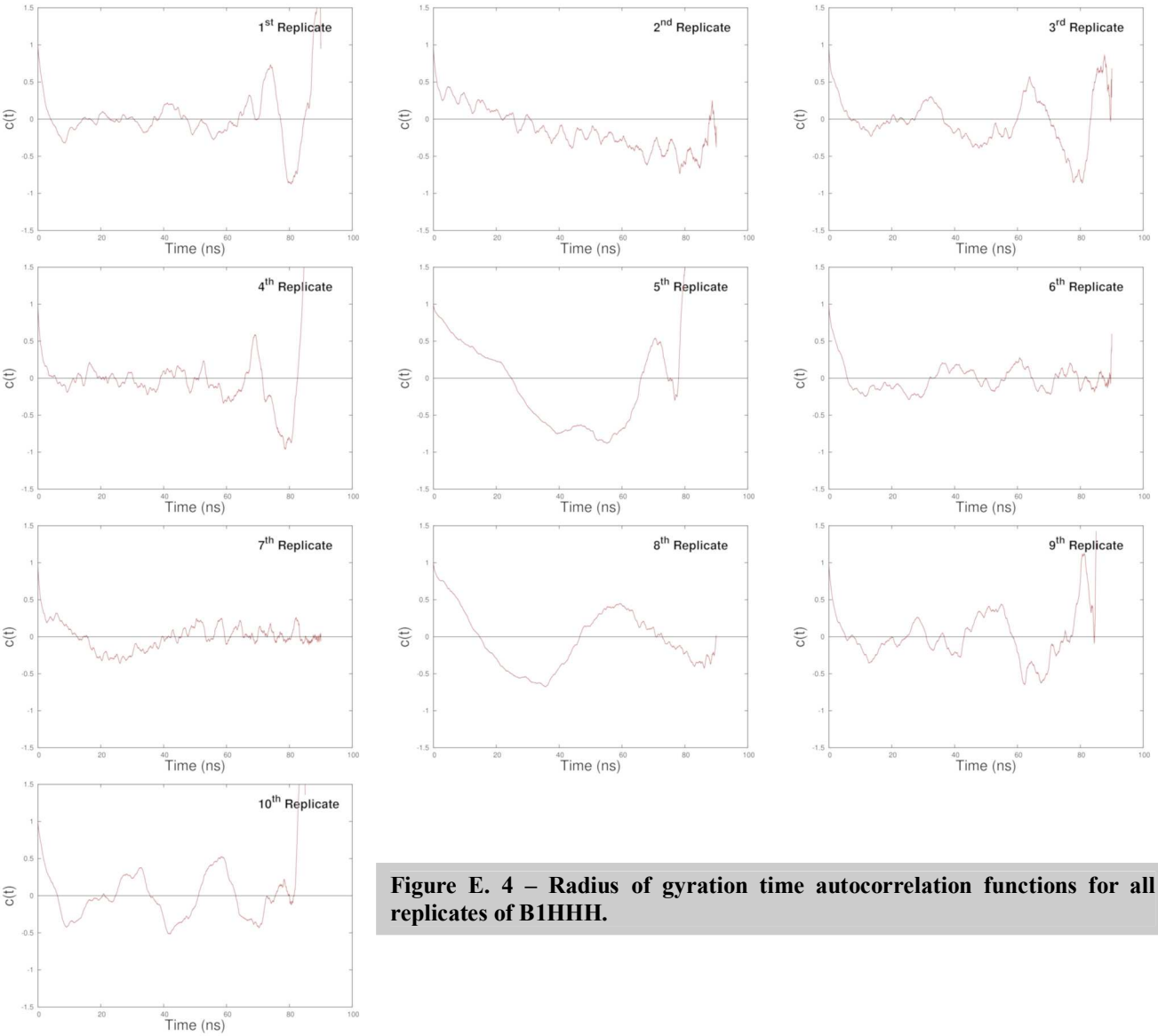


Figure E. 4 – Radius of gyration time autocorrelation functions for all replicates of B1HHH.

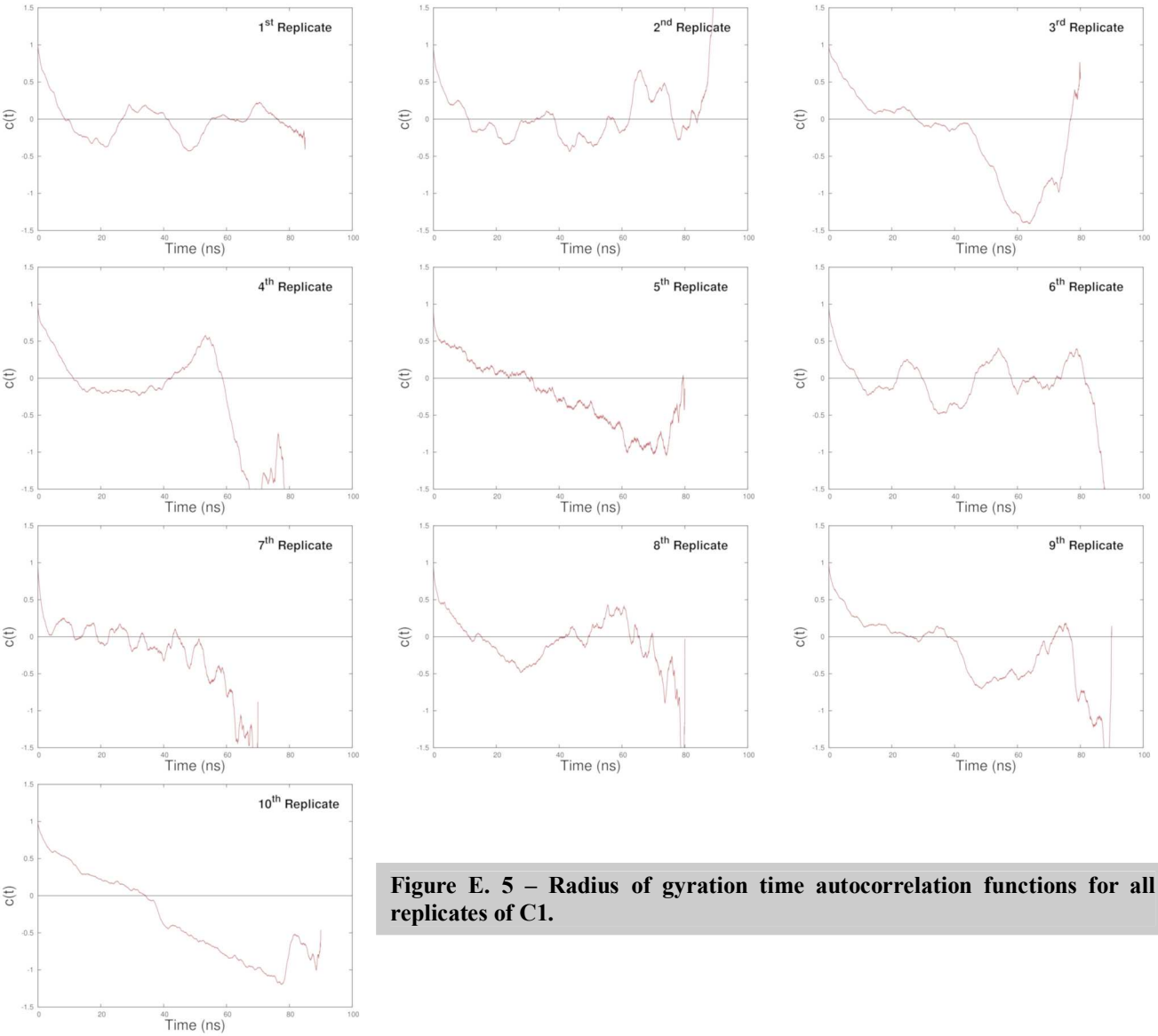
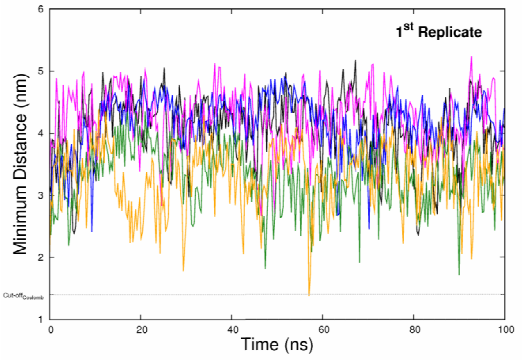
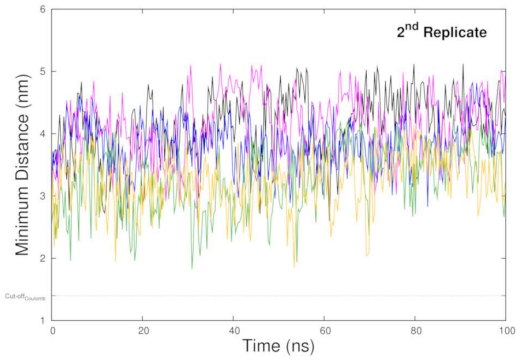


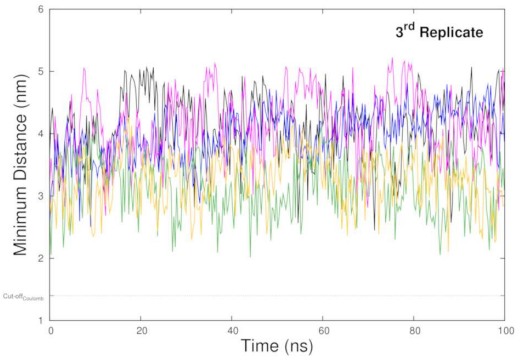
Figure E. 5 – Radius of gyration time autocorrelation functions for all replicates of C1.

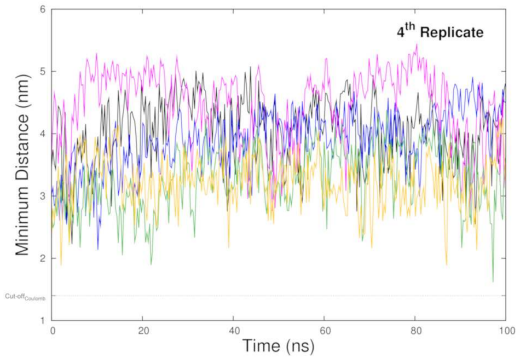
Appendix F

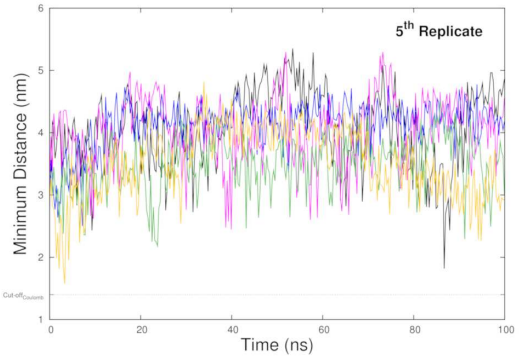
Time series of the minimum distance between periodic images for all dendrimers

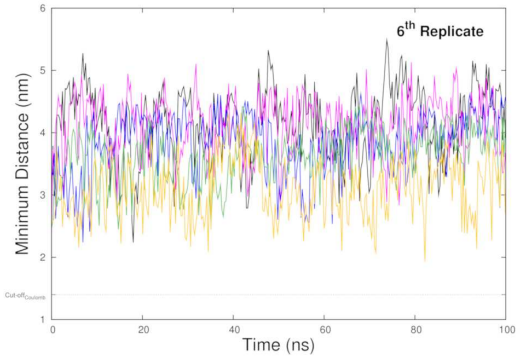












Continues on next page

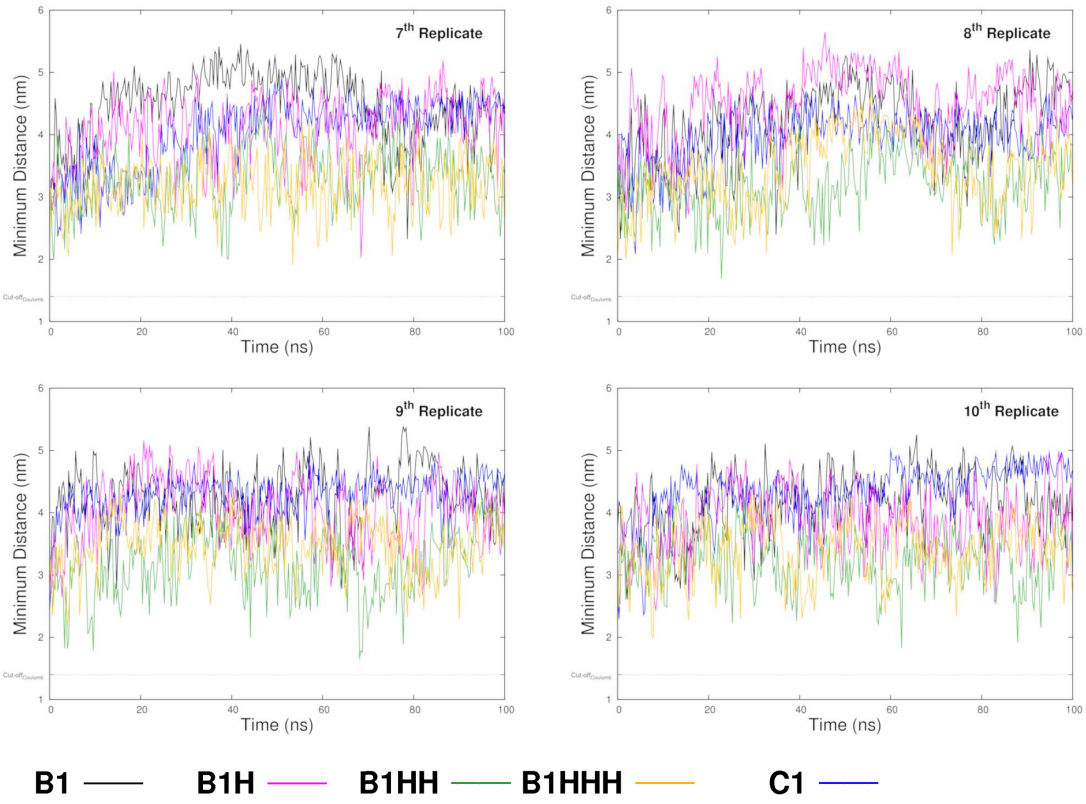
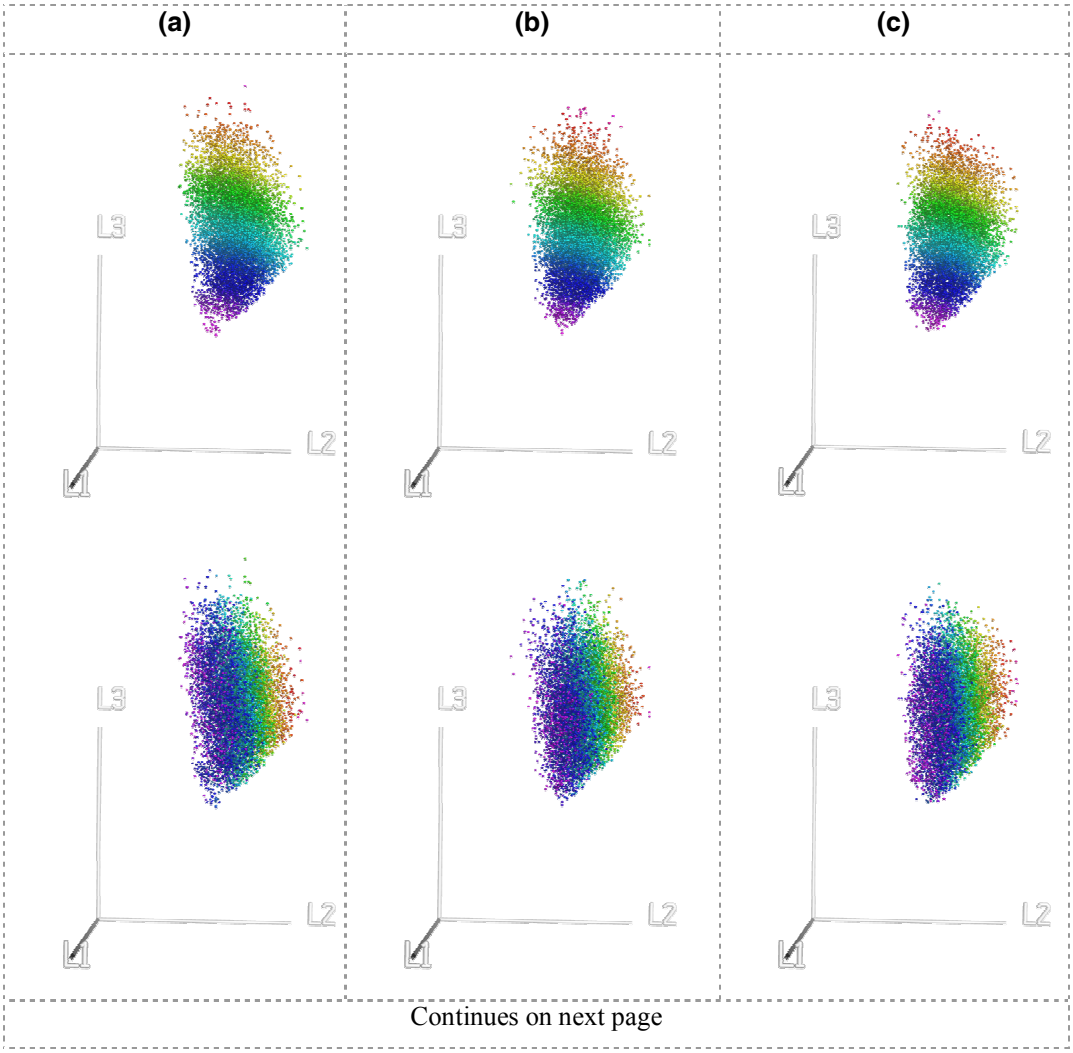


Figure F.1. – Time series of the minimum distance between periodic images for all dendrimers.

Values of the minimum distance between dendrimers periodic images as a function of time for all ten replicates. The 100 ns trajectories are depicted using points collected every 20 picoseconds.

Appendix G

Asphericity, acylindricity and gyration space representation for B1H, B1HH and B1HHH



Continues on next page

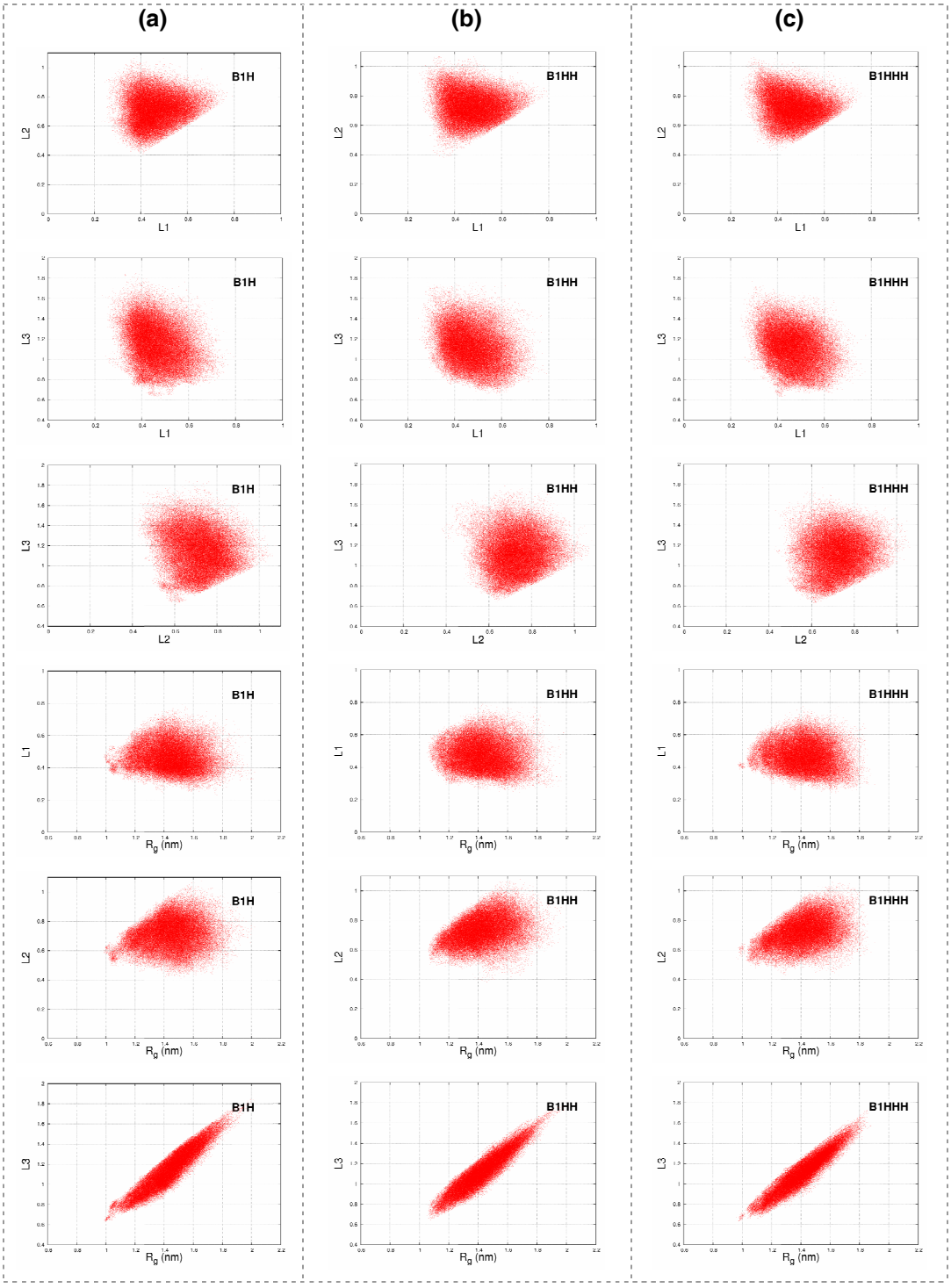


Figure G.1 – Asphericity, acylindricity and gyration space representation for B1H, B1HH and B1HHH

Two identical views of the conformations distribution in gyration space, where the one on the top is colored according to the value of asphericity of each conformation; and the one after, by the acylindricity values. The color scale represents the higher values as red and the lowest as violet. Due to the computational effort involved in constructing the 3D plot we have only represented a tenth of the total amount of points for which we have determined L_1, L_2 and L_3. Relations among the different principal components of the radius of gyration tensor. Relation between the radius of gyration and its diagonalized tensor components. **(a)** B1H; **(b)** B1HH; **(c)** B1HHH.

Appendix H

The concepts behind principal coordinate analysis

In this appendix we present some key ideas collected from works performed using principal coordinate analysis.

Multidimensional scaling (MDS) is a data reduction technique concerned with the problem of finding a set of points in low dimension that represents the "configuration" of the data in high dimension [Cox 2001]. As long as the two entities present a relation among them (regardless of the nature of the relation) that can be described as a (dis)similarity measure, they can be considered as input for MDS.

Principal coordinate analysis (PCoorA) is often taken as synonymous with classical multidimensional scaling, and metric multidimensional scaling, although the latter encompasses more than this one technique [Cox 2001].

PCoorA was developed by Gower in 1966 [Gower 1966]. The starting point for this method is the principal component (PCA) method, which is a statistical technique that linearly transforms an original set of variables into a substantially smaller set of uncorrelated variables representing most of the information in the original dataset [Rencher 2002].

Gower showed the duality between PCoorA and PCA methods. Both methods start from a matrix of dissimilarities between sets of observations and produce a low-dimensional graphical representation of the data, in a such a way that the distances between points in the plot are close to the original, full-dimensional distances which are truly measures of dissimilarity [Steinhauser 2005].

Appendix H

In general, PCA and PCoorA project the $n \times m$ data matrix **M** (a distribution of *n* points in an *m* variable space) on a transformed axes set in which a low-dimensional subspace, containing most of the relational information about the original distribution can be identified [Levy 2001]. In the context of conformational analysis this matrix holds a set of *n* conformations described by the points $P_i(q_{i1}, q_{i2}, \ldots, q_{im})$ in an *m*-dimensional conformation space.

However, while PCA uses the *m*-coordinate vectors **q** associated with the *n* conformation, calculating and operating on the square $m \times m$ $\mathbf{M}^{\mathbf{T}}\mathbf{M}$ matrix (called the covariance matrix, **C**), and thus reflecting the relationship between the *coordinates*, PCoorA performs its operations on the square $n \times n$ $\mathbf{M}\mathbf{M}^{\mathbf{T}}$ matrix, know as the "distance matrix", Δ, therefore, reflecting the relationships between the *conformations* [Levy 2001].

This matrix is transformed into a centered matrix which is then diagonalized. The normalized eigenvectors of the diagonalized matrix are the principal components of the system, and the corresponding eigenvalues indicate the variation along these principal components [Becker 1996; Levy 2001].

A noteworthy observation is that PCoorA-solution is indeterminate with respect to rotation, reflection and shifts [Cox 2001].

A crucial element in PCoorA is the choice of the distance measure used to construct Δ; and it has been suggested that the use of, e.g., Cartesian distances or distances in the dihedral angle space, has a marked effect on the resulting projections [Becker 1996; Elmaci 1999].

In the present work the distance between any two conformations is measured as the RMSD in Cartesian coordinates. The RMSD between all pairs of conformations is determined through Equation 3.6. using all the atoms that compose each conformation.

Appendix H

Classical Example

A classical, but still good example, of what PCoorA does is obtained from the Portuguese geography. Suppose one is confronted with a map from Portugal, and asked to measure (using the metric system), the distances between six Portuguese cities. Then, building a matrix containing all relative distances between cities is a straightforward tasks.

However, let us now reverse the problem; given a matrix with the relative distances between all six Portuguese cities, can one find the original positions of the cities and recreate the map (assuming Portugal is a two dimensional euclidean plane)?

This is a far more difficult task. Still the answer is yes. With the help of a ruler and a compass the problem can even be analytically solved in 2D. With few data points one can even do (2D) PCoorA by hand. To find a PCoorA-solution for the data, first cut six small pieces of paper and write in each of them a different city name (from the set in the figure). Place these pieces of paper arbitrarily in a plane and then move them around in small steps so that higher correlations tend to correspond to smaller distances. Repeat these corrective point movements a few times until the match of distances is satisfactory or until it cannot be improved any more.

This analytical approach is easy for a small number of variables. For many variables, computer algorithms are needed. Therefore, PCoorA is a statistical approach for solving this reverse problem in arbitrary dimensions.

Bibliography

Allen, M. P.; Tildesley, D.J.; "Computer Simulation of Liquids", *Oxford University Press*, 1989, New York.

Ballauff, M.; Likos, C.N.; "Dendrimers in Solution: Insight from Theory and Simulation", *Angew. Chem. Int. Ed.*, 2004, 43:2998-3020.

Bauer, R.E.; Enkelmann, V.; Wiesler, U.M.; Berresheim, A.J.; Müllen, K.; "Single-Crystal Structures of Polyphenylene Dendrimers", *Chem.- Eur. J.*, 2002, 8:3858-3864.

Becker, O.M.; "Principal Coordinate Maps of Molecular Potential Energy Surfaces", *J. Comp. Chem.*, 1996, 19:1255-1267.

(a) Becker, O.M.; Karplus, M.; "The topology of multidimensional potential energy surfaces: Theory and application to peptide structure and kinetics", *J. Chem. Phys.*, 1997, 106:1495-1517.

(b) Becker, O.M.; "Geometrical versus topological clustering: An insight into conformation mapping", *Proteins*, 1997, 27:213–226.

Becker, U.M.; MacKerrel Jr., A.D.; Roux, B.; Watanabe, M.; "Computational Biochemistry and Biophysics", *Marcel Dekker Inc.*, 2001, New York.

Berendsen, H.C.; van der Spoel, D.; van Drunen, R.; "GROMACS: A message-passing parallel molecular dynamics implementation", *Comp. Phys. Comm.*, 1995, 91:43–56.

Berendsen, H.J.C.; "Simulating the Physical World", *Cambridge University Press*; 2007, New York.

Berendsen, H.J.C.; Postma, J.P.M.; van Gusteren, W.F.; DiNola, A.; Haak, J.R.; "Molecular dynamics with coupling to an external bath", *J. Chem. Phys.*, 1984, 81:3684-3690.

Berendsen, H.J.C.; Postma, J.P.M.; van Gusteren, W.F.; Hermans, J.; "Interaction Models for Water in Relation to Protein Hydration", *Intermolecular Forces. Pullman, B. (ed)*, 1981, D. Reidel Publishing Company, Dordrecht, 331–342.

Bezouska, K.; "Design, functional evaluation and biomedical applications of carbohydrate dendrimers (glycodendrimers)", *J. Biotechnol.*, 2002, 90:269-290.

Boas, U.; Heegaard, P.M.; "Dendrimers in drug research", *Chem. Soc. Rev.*, 2004, 33:43-63.

Bosman, A.W.; Janssen, H.M.; Meijer, E.W.; "About Dendrimers: Structure, Physical Properties, and Applications", *Chem. Rev.*, 1999, 99:1665-1688.

Bracci, L.; 0Falciani, C.; Lelli, B.; Lozzi, L.; Runci, Y.; Pini, A.; De Montis, M.G.; Tagliamonte, A.; Neri, P.; "Synthetic Peptides in the Form of Dendrimers Become Resistant to Protease Activity", *J. Biol. Chem.*, 2003, 278:46590–46595.

Breslow, R.; Wei, S.; Kenesky, C.; "Enantioselective transaminations by dendrimeric enzyme mimics", *Tetrahedron*, 2007, 63:6317-6321.

Brooks, C.L.; "Protein and Peptide Folding Explored with Molecular Simulations", *Acc. Chem. Res.*, 2002, 35:447-454.

Brunner, H.; "Dendrizymes: Expanded ligands for enantioselective catalysis", *J. Organomet. Chem*, 1995, 500:39-46.

Buhleier, E.W.; Wehner, W.; Vögtle, F.; “"Cascade"- and "Nonskid-Chain-like" Syntheses of Molecular Cavity Topologies”, *Synthesis*, 1978, 1978:155-158.

Çagın, T.; Wang, G.; Martin, R.; Breen, N.; Goddard III, W.A.; “Molecular modelling of dendrimers for nanoscale applications”, *Nanotechnology*, 2000, 11:77–84.

Campos, S.R.R.; Baptista, A.M.; “Conformational analysis in a multidimensional energy landscape: study of an arginylglutamate repeat”, *J. Phys. Chem. B*, 2009, Accepted for Print. (DOI: 10.1021/jp902991u)

Cantor, C.R.; Schimmel, P.R.; “Biophysical Chemistry”, *W. H. Freeman And Company*, 1980; New York.

Chandler, D.;“Introduction to Modern Statistical Mechanics”, *Oxford University Press*, 1987, New York.

Chen, Z.Y.; Cui, S.-M.; “Monte Carlo Simulations of Star-Burst Dendrimers”; *Macromolecules*, 1996, 29:7943-7952.

Chitra, R.; Yashonath, R.; “Estimation of Error in the Diffusion Coefficient from Molecular Dynamics Simulations”, *J. Phys. Chem. B*, 1997, 101:5437-5445.

Cloninger, M.J.; “Biological applications of dendrimers”, *Curr. Op. Chem. Biol.*, 2002, 6:742-748.

(a) Clouet, A.; Darbre, T.; Reymond, J.; “A Combinatorial Approach to Catalytic Peptide Dendrimers”, *Angew. Chem. Int. Ed.*, 2004, 43:4612-4615.

(b) Clouet, A.; Darbre, T.; Reymond, J.-L.; “Esterolytic Peptide Dendrimers with a Hydrophobic Core and Catalytic Residues at the Surface”, *Adv. Synth. Catal.*, 2004, 346:1195-1204.

Cohen, Y.; Avram, L.; Frish, L.; “Diffusion NMR Spectroscopy in Supramolecular and Combinatorial Chemistry: An Old Parameter—New Insights”, *Angew. Chem. Int. Ed.*, 2005, 44:520-554.

Cox, T.F.; Cox, M.A.A.; “Multidimensional Scaling. Monographs on Statistic and Applied Probability”, 2nd edition, *Chapman & Hall/CRC*, 2001, Boca Raton.

Crespo, L.; Sanclimens, G.; Pons, M.; Giralt, E.; Royo, M.; Albericio, F.; “Peptide and amide bond-containing dendrimers”, *Chem. Rev.*, 2005, 105:1663-1681.

Darbre, T.; Reymond, J.; “Peptide Dendrimers as Artificial Enzymes, Receptors, and Drug-Delivery Agents”, *Acc. Chem. Res.*, 2006, 39:925–934.

Daura, X.; Gademann, K.; Jaun, B.; Seebach, D.; van Gunsteren, W.F.; Mark, A.E.; “Peptide Folding: When Simulation Meets Experiment”, *Angew. Chem. Int. Ed.*; 1999, 38:236-240.

DeLano, W. L., “The PyMOL Molecular Graphics System. Palo Alto”, *DeLano Scientific*, 2002, CA, USA.(http://pymol.org/)

Delort, E.; Nguyen-Trung, N.; Darbre, T.; Reymond, J.-L.; “Synthesis and Activity of Histidine-Containing Catalytic Peptide Dendrimers”, *J. Org. Chem.*, 2006, 71:4468-4480.

Diederich, F.; Felber, B.; “Supramolecular chemistry of dendrimers with functional cores”, *Proc. Nat. Aca. Sci.*, 2002, 99:4778-4781.

Douat-Casassus, C.; Darbre, T.; Reymond J.-L.; “Selective Catalysis with Peptide Dendrimers”, *J. Am. Chem. Soc.*, 2004, 126:7817-7826.

Dykes, G.M.; “Dendrimers: a review of their appeal and applications”, *J. Chem. Technol. Biotechnol.*, 2001, 76:903-918.

Elmaci, N.; Berry, R.S.; “Principal coordinate analysis on a protein model”, *J. Chem. Phys.*, 1999, 110:10606-10622.

Esposito, A.; Delort, E.; Lagnoux, D.; Djojo, F.; Reymond, J.-L.; “Catalytic Peptide Dendrimers”, *Angew. Chem. Int. Ed.*, 2003, 42:1381-1383.

Everitt, B.S.; Landau, S.; Leese, M.; “Cluster Analysis”, 4th ed., *Oxford University Press*, 2001, New York.

Flory, P.J.; "Molecular Size Distribution in Three Dimensional Polymers. I. Gelation", *J. Am. Chem. Soc.*, 1941, 63:3083-3090.

Frauenfelder, H.; Leeson, D.T.; "The energy landscape in non-biological and biological molecules", *Nature*, 1998, 5:757-759.

Fréchet, J.M.; "Dendrimers and supramolecular chemistry", *Proc. Natl. Am. Soc.*, 2002, 17:117-132.

Freire, J.J.; Rubio, A.M.; "Conformational properties and Rouse dynamics of different dendrimer molecules", *Polymer*, 2008, 49:2762-2769.

Friedhofen, J.H.; Vögtle, F.; "Detailed nomenclature for dendritic molecules", *New J. Chem.*, 2006, 30:32-43.

Gennes, P.G.; Hervet, H.; "Statistics of starburst polymers", *J. Phys. Lett.*, 1983, 44:L351-L360.

Genzer, J.; Kolafa, J.; "Molecular dynamics of potential models with polarizability: comparison of methods", *J. Mol. Liq.*, 2004, 109:63-72.

Gillies, E.R.; Fréchet, J.M.; "Dendrimers and dendritic polymers in drug delivery", *Drug Discovery Today*, 2005, 10:35-43.

Gillies, E.R.; Fréchet, J.M.; "Designing Macromolecules for Therapeutic Applications: Polyester Dendrimers Poly(ethylene oxide) "Bow-Tie" Hybrids with Tunable Molecular Weight and Architecture", *J. Am. Chem. Soc.*, 2002, 124, 14137.

Goldstein, H.; "Classical Mechanics", 2th edition, *Addison-Wesley Publishing Company*, 1980, U.S.A.

Gorman, C.B.; Smith, J.C.; "Structure – Property Relationships in Dendritic Encapsulation", *Acc. Chem. Res.*, 2001, 34:60-71.

Götze, I.O.; Likos, C.N.; "Conformations of Flexible Dendrimers: A Simulation Study", *Macromolecules*, 2003, 36:8189-8197.

Gower, J.C.; "Some distance properties of latent root and vector methods used in multivariant analysis", *Biometrika*, 1966, 53:325–338.

Grinstaff, M.W.; "Biodendrimers: New Polymeric Biomaterials for Tissue Engineering", *Chem. Eur. J.*, 2002, 8:2838-2846.

Guo, Z.; Brooks III, C.L.; Boczko, E.M.; "Exploring the folding free energy surface of a three-helixbundle protein", *Proc. Natl. Acad. Sci.*, 1997, 94:10161-10166.

Hamprecht, F.A.; Peter, C.; Daura, X.; Thiel, W.; van Gusteren, W.F.; "A strategy for analysis of (molecular) equilibrium simulations: Configuration space density estimation, clustering, and visualization", *J. Chem. Phys.*, 2001, 114:2079-2089.

Han, M.; Chen, P.; Yang, X.; "Molecular dynamics simulation of PAMAM dendrimer in aqueous solution", *Polymer*, 2005, 46:3481-3488.

Hawker, C.J.; Fréchet, J.M.; "Preparation of Polymers with Controlled Molecular Architecture. A New Convergent Approach to Dendritic Macromolecules", *J. Am. Chem. Soc.*, 1990, 112:7638-7647.

(a) Hawker, C.J.; Wooley, K.L.; Fréchet, J.M.; "Solvatochromism as a Probe of the Microenvironment in Dendritic Polyethers: Transition from an Extended to a GlobuIar Structure", *J. Am. Chem. Soc.*, 1993, 115:4375-4376.

(b) Hawker, C.J.; Wooley, K.L.; Fréchet, J.M.; "Unimolecular Micelles and Globular Amphiphiles: Dendritic Macromolecules as Novel Recyclable Solubilization Agents", *J. Chem. Soc. Perkin Trans.*, 1993, 1:1287-1297.

Hecht, S.; Fréchet, J.M.; "Dendritic Encapsulation of Function: Applying Nature's Site Isolation Principle from Biomimetics to Materials Science", *Angew. Chem. Int. Ed. Engl.*, 2001, 40:74-91.

Hess, B.; "Stochastic Concepts in Molecular Simulation, Diss", 2002, University of Groninger.

Hess, B.; Bekker, H.; Berendsen, H.J.C.; Fraaije J.G.E.M.; "LINCS: A Linear Constraint Solver for Molecular Simulations". *Journal of Computational Chemistry*, 1997, 18:1463–1472.

Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E.; "GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation", *J. Chem. Theory Comput.*, 2008, 4:435-447.

Hess, B.; van der Vegt, N.F.A.; "Hydration Thermodynamic Properties of Amino Acid Analogues: A Systematic Comparison of Biomolecular Force Fields and Water Models", *J. Phys. Chem. B*, 2006, 110:17616-17626.

Hovmöller, S.; Zhou, T.; Ohlson, T.; "Conformations of amino acids in proteins", *Acta Cryst.*, 2002, D58:768-776.

Hünenberger, P.H.; "Molecular dynamics simulations using empirical force fields: Principles and applications to selected systems of chemical and biochemical interest", *PhD Thesis*, 1997, Zürich: ETH Zürich.

Jansen, J.S.; de Brabander-van den Berg, E.M.; "Encapsulation of Guest Molecules into a Dendritic Box", *Science*, 1994, 266:1226-1229.

Javor, S.; Nataello, A.; Doglia, S.M.; Reymond, J.; "α-Helix Stabilizatopn within a Peptide Dendrimer", *J. Am. Chem. Soc.*, 2008, 130:17248-17249.

Javor, S.; Reymond, J.; "Molecular Dynamics and Docking Studies of Single Site Esterase Peptide Dendrimers", *J. Org. Chem.*, 2009, 74:3665–3674.

Javor, S; Delort, E.; Darbre, T.; Reymond, J.-L.; "A Peptide Dendrimer Enzyme Model with a Single Catalytic Site at the Core", *J. Am. Chem. Soc.*, 2007, 129:13238-13246.

Johansson, E.M.; Crusz, S.A.; Kolomiets, E.; Buts, L.; Kadam, R.U.; Cacciarini, M.; Bartels, K.; Diggle, C.P.; Cámara, M.; Williams, P.; Loris, R.; Nativi, C.; Rosenau, F.; Jaeger, K.; Darbre, T.; Reymond, J.-L.; "Inhibition and Dispersion of Pseudomonas aeruginosa Biofilms by Glycopeptide Dendrimers Targeting the Fucose-Specific Lectin LecB", *Chem. Biol.*, 2008, 15:1249-1257.

Johansson, E.M.; Kolomiets, E.; Rosenau, F.; Jaeger, K.; Darbre, T.; Reymond, J.-L.; "Combinatorial variation of branching length and multivalency in a large (390 625 member) glycopeptide dendrimer library: ligands for fucose-specific lectins", *New J. Chem.*, 2007, 31:1291-1299.

Karatasos, K.; Adolf, D.B.; Davies, G.R.; "Statics and dynamics of model dendrimers as studied by molecular dynamics simulations", *J. Chem. Phys.*, 2001, 115:5310-5318.

Karplus, M.; McCammon, J.A.; "Molecular dynamics simulations of biomolecules", *Nature Struc. Biol.*, 2002, 9:646-652.

Kim, Y.; Zeng, F.; Zimmerman, S.C.; "Peptide Dendrimers from Natural Amino Acids", *Chem. Eur. J.*, 1999, 5:2133-2138.

Kleij, A.W.; Gossage, R.A.; Gebbink, R.K.; Brinkmann, N.; Reijerse, E.J.; Kragl, U.; Lutz, M.; Spek, A.L.; van Koten, G.; "A 'Dendritic Effect' in Homogeneous Catalysis with Carbosilane-Supported Arylnickel(II) Catalysts: Observation of Active-Site Proximity Effects in Atom-Transfer Radical Addition", *J. Am. Chem. Soc.*, 2000, 112:12112-12124.

Klepeis, J.L.; Lindorff-Larsen, K.; Dror, R.; Shaw, D.E.; "Long-timescale molecular dynamics simulations of protein structure and function", *Curr. Op. Chem. Biol.*, 2009, 19:120-127.

Kleywegt, G.J.; Jones, T.A.; "Phi/Psi-chology: Ramachandran revisited", *Structure*, 1996, 4:1395-1400.

Klos, J.S.; Sommer, J.-U.; "Properties od Dendrimers with Flexible Spacer-Chains: A Monte Carlo Study", *Macromolecules*, 2009, 42:4878-4886.

Kono, K.; Fukui, T.; Takagishi, T.; Sakurai, S.; Kojima, C.; "Preparation of poly(ethylene glycol)-modified poly(amidoamine) dendrimers with a shell of hydrophobic amino acid residues and their function as a nanocontainer", *Polymer*, 2008, 49:2832-2838.

Laidler, K.J.; "A glossary of terms used in chemical kinetics, including reaction dynamics (IUPAC Recommendations 1996)", *Pure Appl. Chem.*, 1996, 68:149-192.

Lazaridis, T.; Karplus, M.; ""New View" of Protein Folding Reconciled with the Old Through Multiple Unfolding Simulations", *Science*, 1997, 278:1928-1931.

Leach, A.R.; "Molecular Modeling: Principles and Applications",2th ed., *Pearson Education Limited*, 2001, London.

Lee, H.; Baker Jr., J.R.; Larson, R.G.; "molecular dynamics studies of the size, shape and internal structure of 0% and 90%-acetylated G5 PAMAM dendrimers in water and methanol", *J. Phys. Chem. B*, 2006, 110:4014-4019.

Lee, I.; Athey, B.D.; Wetzel, A.W.; Meixner, W.; Baker, J.R.; "Structural Molecular Dynamics Studies on Polyamidoamine Dendrimers for a Therapeutic Application: Effects of pH and Generation", *Macromolecules*, 2002, 35:4510-4520.

Lescanec, R.L.; Muthukumar, M.; "Configurational characteristics and scaling behavior of starburst molecules: a computational study", *Macromolecules*, 1990, 23:2280-2288.

(a) Levy, Y.; Becker, O.M.; "Energy landscapes of conformationally constrained peptides", *J. Chem. Phys*, 2001, 114:993-1009.

(b) Levy, Y.; Jortner, J.; Becker, O.M.; "Dynamics of hierarchical folding on energy landscapes of hexapeptides", *J. Chem. Phys.*, 2001, 115:10533-10547.

Lin, S.; Maiti, P.K.; Goddard III, W.A.; "Dynamics and Thermodynamics of Water in PAMAM Dendrimers at Subnanosecond Time Scales", *J. Phys. Chem. B*, 2005, 109:8663-8672.

Lindahl, E.; Hess, B.; van der Spoel, D.; "Gromacs 3.0: A package for molecular simulation and trajectory analysis", *J. Mol. Mod.*, 2001, 7:306–317.

Liu, M.J.; Fréchet, J.M.; "Designing dendrimers for drug delivery", *Pharm. Sci. Technol. Today*, 1999, 2:393-401.

Lockman, J.W.; Paul, N.M.; Parquette, J.R.; "The role of dynamically correlated conformational equilibria in the folding of macromolecular structures. A model for the design of folded dendrimers", *Prog. Polym. Sci.*, 2005, 30:423–452.

Lovell, S.C.; Davis, I.W.; Arendall III, W.B.; Bakker, P.I.W.; Word, J.M.; Prisant, M.G.; Richardson, J.S.; Richardson, D.C.; "Structure Validation by Cα Geometry: φ,ψ and Cβ Deviation", *Proteins*, 2003, 50:437-450.

Lyulin, A.V.; Davies, G.R.; Adolf, D.B.; "Location of Terminal Groups of Dendrimers: Brownian Dynamics Simulation"; *Macromolecules*, 2000, 33:6899-6900.

Machuqueiro, M.; Baptista, A.M.; "Constant-pH Molecular Dynamics with Ionic Strength Effects: Protonation-Conformation Coupling in Decalysine", *J. Chem. Phys. B.*, 2006, 110:2927-2933.

Maillard, N.; Clouet, A.; Darbre T.; Reymond, J.; "Combinatorial libraries of peptide dendrimers: design, synthesis, on-bead high-throughput screening, bead decoding and characterization", *Nature Protocols*, 2009, 4:132-142.

Maisuradze, G.G.; Liwo, A.; Scheraga, H.A.; "Principal Component Analysis for Protein Folding Dynamics", *J. Mol. Biol.*, 2009, 385:312–329.

Maiti, P.K.; Çağin, T.; Lin, S.; Goddard III, W.A.; "Effect of Solvent and pH on the Structure of PAMAM Dendrimers", *Macromolecules*, 2005, 38:979-991.

Maiti, P.K.; Çağin, T.; Wang, G.; Goddard III, W.A.; "Structure of PAMAM Dendrimers: Generation 1 through 11", *Macromolecules*, 2004, 37:6236-6254.

Mansfield, M.L.; Jeong, M.; "Simulation of Lattice Dendrimers by a Monte Carlo Technique with Detailed Balance", *Macromolecules*, 2002, 35:9794-9798.

Mattice, W.L.; Suter, U.W.; "Conformational Theory of Large Molecules", *Wiley-Interscience*, 1994, New York.

Mattice,w.L.; "Persistence Vectors, Average Center of Mass Vectors, and Moment of Inertia Tensors for Branched Polymethylenes", *Macromolecules*, 1997, 10:1182-1186.

Mazur, J.; Guttman, C.M.; McCrackin, F.L.; "Monte Carlo Studies of Self-Interacting Polymer Chains with Excluded Volume. 11. Shape of a Chain", *Macromolecules*, 1973, 6:872-874.

Medina, S.H.; El-Sayed, M.H.; "Dendrimers as Carriers for Delivery of Chemotherapeutic Agents", *Chem. Rev.*, 2009, 109:3141–3157.

Méry, D.; Astruc, D.; "Dendritic catalysis: Major concepts and recent progress", *Coordination Chemistry Reviews*, 2006, 250:1965-1979.

Micaelo, N.M.; Baptista, A.M.; Soares, C.M.; "Parametrization of 1-Butyl-3-methylimidazolium Hexafluorophosphate/Nitrate Ionic Liquid for the GROMOS Force Field", *J. Phys. Chem. B*, 2006, 110:14444-14451.

Miyamoto, S., Kollman, P. A. "SETTLE: An analytical version of the SHAKE and RATTLE algorithms for rigid water models.", *J. Comp. Chem.*, 1992, 13:952–962.

Molecular Simulation, ITQB: http://www.itqb.unl.pt/labs/molecular-simulation

Moss, J.P.; "Basic terminology of stereochemistry (IUPAC Recommendations 1996)", Pure Appl. Chem., 1996, 68:2193-2222.

Murat, M.; Grest, G.S.; "Molecular Dynamics Study of Dendrimer Molecules in Solvents of Varying Quality", *Macromolecules*, 1996, 29:1278-1285.

Narayanan, V.V.; Newkome, G.R.; "Supramolecular Chemistry within dendritic structures", *Top Curr Chem*, 1998, 197:19-77.

Naylor, A.M.; Goddard III, W.A.; Kiefer, G.E.; Tomalia, A.D.; "Starburst Dendrimers. 5. Molecular Shape Control", *J. Am. Chem. Soc.*, 1989, 111:2339-2341.

Newkome, G.R.; Yao, Z.; Baker, G.R.; Gupta, V.K.; "Cascade Molecules: A New Approach to Micelles", *J. Org. Chem.*, 1985, 50:2003-2004.

Niedrhafner, P.; Šebestík, J.; Ježek, J.; "Peptide dendrimers", *J. Peptide Sci.*, 2005, 11:757-788.

Oliveira, A.S.O.; Teixeira, v.H.; Baptista, A.M.; Soares, C.M.; "Reorganization and Conformational Changes in the Reduction of Tetraheme Cytochromes", *Biophysical J.*, 2005, 89:3919-3930.

Oostenbrink, C.; Soares, T.A.; van der Vegt, N.F.A.; van Gunsteren, W.F.; "Validation of the 53A6 GROMOS force feild", *Eur. Biophys. J.*, 2005, 34:273-284.

Oostenbrink, C.; Villa, A.; Mark, A.E.; van Gunsteren, W.F.; "A Biomolecular Force Field Based on the Free Enthalpy of Hydration and Solvation: The GROMOS Force-Field Parameter Sets 53A5 and 53A6", *J. Comp. Chem.*, 2004, 25:1656-1676.

Patri, A.K.; Majoros, I.J.; Baker, J.R.; "Dendritic polymer macromolecular carriers for drug delivery",*Curr. Op. Chem. Biol.*, 2002, 6:466-471.

Plotkin, S.S.; Onuchic, J.N.; "Understanding protein folding with energy landscape theory. Part I: Basic concepts", *Quaternary Reviews of Biophysics*, 2002, 35:111-167.

Ponder, J.W.; Case, D.A.; "Force Fields for Protein Simulations", *Adv. Prot. Chem.*, 2003, 66:27-85.

Protein Modeling, ITQB: http://www.itqb.unl.pt/labs/protein-modelling

Ranganathan, D.; Kurur, S.; "Synthesis of Totally Chiral, Multiple Armed, Poly Glu and Poly Asp Scaffoldings on Bifunctional Adamantane Core", *Tetrahedron Lett.*, 1997, 38:1265-1268.

Rapaport, D.C.; "The Art of Molecular Dynamics Simulations", 2th ed., *Cambridge University Press*, 2004, New York.

Rencher, A.C.; "Methods of Multivariate Analysis", 2th edition, *Wiley-Interscience*, 2002, New York.

Robertson, A.; Luttmann, E.; Pande, V.S.; "Effects of Long-Range Electrostatic Forces on Simulated Protein Folding Kinetics", *J. Comput. Chem.*, 2008, 29:694-700.

Sadler, K.; Tam, J.T.; "Peptide dendrimers: applications and synthesis", *Rev. Mol. Biotechnol.*, 2002, 90:195-229.

Shao, J.; Tanner, S.W.; Thompsdon, N.; Cheatham III, T.E.; "Clustering Molecular Dynamics Trajectories: 1. Characterizing the Performance of Different Clustering Algorithms", *J. Chem. Theory Comput.*, 2007, 3:2312-2334.

Shea, J.-E.; Brooks III, C.L.; "From Folding Theories to Folding Proteins, A Review and

Assessment of Simulation Studies of Protein Folding and Unfolding", *Annu. Rev. Phys. Chem.*, 2001, 52:499-535.

Shi, X.; Wang, S.H.; Lee, I.; Shen, M.; Baker Jr., J.R.; "Comparison of the Internalization of Targeted Dendrimers and Dendrimer-Entrapped Gold Nanoparticles into Cancer Cells", *Biopolymer*, 2009, Accepted for publication.

Smith, D.K.; Diederich, F.; "Functional Dendrimers: Unique Biological Mimics", *Chem. Eur. J.*, 1998, 4:1353-1361.

Smith, P.E.; van Gunsteren, W.F.; "Consistent dielectric properties of the simple point charge and extended simple point charge water models at 277 and 300K", *J. Chem. Phys.*, 1994, 100:3169-3174.

Šolc, K.; "Statistical Mechanics of Random-Flight Chains. IV. Size and Shape Parameters of Cyclic, Star-like, and Comb-like Chains", *Macromolecules*, 1973, 6:378-385.

Sommer, P.; Fluxa, V.S.; Darbre, T.; Reymond, J.; "Proteolysis of Peptide Dendrimers", *Che. Bio. Chem.*, 2009, 10:1527-1536.

Sommer, P.; Uhlich, N.A.; Reymond, J.-L.; Darbre, T.; "A Peptide Dendrimer Model for Vitamin B12 Transport Proteins", *ChemBioChem.*, 2008, 9:689–693.

Steinhauser, M.O.; "A molecular dynamics study on universal properties of polymer chains in different solvent qualities. Part I. A review of linear chain properties", *J. Chem. Phys.*, 2005, 122:094901.

Stevelmans, S.; van Hest, J.C.; Jansen, J.F.; van Boxtel, D.A.; de Brabander-van den Berg, E.M.; Meijer, E.S.; "Synthesis, Characterization, and Guest-Host Properties of Inverted Unimolecular Dendritic Micelles", *J. Am. Chem. Soc.*, 1996, 118:7398-7399.

Stiriba, S.; Frey, H.; Haag, R.; "Dendritic Polymers in Biomedical Applications: From Potential to Clinical Use in Diagnostics and Therapy", *Angew. Chem. Int. Ed.*, 2002, 41:1329-1334.

Tam, J.P.; "Synthetic peptide vaccine design: Synthesis and properties of a high-density multiple antigenic peptide system", *Proc. Natl. Acad. Sci.*, 1988, 85:5409-5413.

Tanis, I.; Tragoudaras, D.; Karatasos, K.; Anastasiadis, S.H.; "Molecular Dynamics Simulations of a Hyperbranched Poly(ester amide): Statics, Dynamics, and Hydrogen Bonding", *J. Phys. Chem. B*, 2009, 113:5356-5368.

Theodoru, D.N.; Suter, U.W.; "Shape of unperturbed polymers: polypropylene", *Macromolecules*, 1985, 18:1206-1214.

Tironi, I.G.; Sperb, R.; Smith, P.E.; van Gunsteren, W.F.; "A generalized reaction field method for molecular dynamics simulations", *J. Chem. Phys.*, 1995, 102:5451-5459.

Tomalia, D. A.; Baker, H.; Dewald, J.; Hall, J. M.; Kallos, G.; Martin, R.; Ryder, J.; "A New Class of Polymers: Starburst-Dendritic Macromolecules", *Polymer*, 1985, 17:117-132.

Tomalia, D.A.; Fréchet, J.M.; "Discovery of Dendrimers and Dendritic Polymers: A Brief Historical Perspective", *J. Polym. Sci. Part A: Polym Chem.*, 2002, 40:2719-2728.

Turnbull , W.B.; Stoddart, J.F.; "Design and synthesis of glycodendrimers", *Rev. Mol. Biotechnol.*, 2002, 90:231-255.

Udgaonkar, J.B.; "Multiple Routes and Structural Heterogeneity in Protein Folding", *Annu. Rev. Biophys.*, 2008, 37:489-510.

van der Spoel, D.; Berendsen, H.J.C, "Molecular Dynamics Simulations of Leu-Enkephalin in Water and DMSO", *Biophysical J.*, 1997, 72:2032-2041.

(a) van der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A.E.; Berendsen, H.C.; "GROMACS: Fast, Flexible and Free", *J. Comp. Chem.*, 2005, 26:1701–1718.

(b) van der Spoel, E. Lindahl, B. Hess, A. R. van Buuren, E. Apol, P. J. Meulenhoff, D. P. Tieleman, A. L. T. M. Sijbers, K. A. Feenstra, R. van Drunen and H. J. C. Berendsen, "Gromacs User Manual version 4.0", www.gromacs.org, (2005).

(a) van der Spoel, D.; van Maaren, P.J.; Berendsen, H.J.C.; "A systematic study of water models for molecular simulation: Derivation of water models optimized for use with a reaction field", *J.*

Chem. Phys., 1998, 108:10220-10230.

(b) van der Spoel, D.; Vogel, H.J.; Berendsen, H.J.C.; "Molecular Dynamics Simulations of N-Terminal Peptides from a nucleotide Binding Protein", *Proteins*, 1998, 24:450-466.

van der Spoel, D.; van Maaren, P.J.; "The Origin of Layer Structure Artifacts in Simulations of Liquid Water", *J. Chem. Theory Comput.*, 2006, 2:1-11.

van Gunsteren, W.F.; Bakowies, D.; Baron, R.; Chandrasekhar, I.; Christen, M.; Daura, X.; Gee, P.; Geerke, D.P.; Glättli, A.; Hünenberger, P.H.; Kastenholz, M.A.; Oostenbrink, M.; Schenk, M.; Trzesniak, D.; van der Vegt, N.F.A.; Yu, H.B.; "Biomolecular Modeling: Goals, Problems, Perspectives", *Angew. Chem. Int. Ed.*, 2006, 45:4064-4092.

Van Gunsteren, W.F.; Berendsen, H.J.C.; "Computer Simultion of Molecular Dynamics: Methodology, Applications, and Perspectives in Chemistry", *Angew. Chem. Int. Ed. Engl.*, 1990, 29:992-1023.

van Gunsteren, W.F.; Hünenberger, P.H.; Mark, A.E.; "Validation of molecular dynamics simulation", *J. Chem. Phys.*, 1998, 108:6109-6116.

van Gunsteren, W.F.; Hünenberger, P.H.; Mark, A.E.; Smith, P.E.; Tironi, I.G.; "Computer simulation of protein motion", *Comput. Phys. Commun.*, 1995, 91:305-319.

Vassura, M.; Margara, L.; Di Lena, P.; Medri, F.; Fariselli, P.; Casadio, R.; "Reconstruction of 3D Structures from Protein Contact Maps", *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 2008, 5:357-367.

(a) Vepřek, P.; Ježek, J.; "Peptide and Glycopeptide Dendrimers. Part I", *J. Peptide Sci.*, 1999, 5:5-23.

(b) Vepřek, P.; Ježek, J.; "Peptide and Glycopeptide Dendrimers. Part II", *J. Peptide Sci.*, 1999, 5:203-220.

Weber, W.; Hu1nenberger, P.H.; McCammon, J.A.; "Molecular Dynamics Simulations of a Polyalanine Octapeptide under Ewald Boundary Conditions: Influence of Artificial Periodicity on Peptide Conformation", *J. Phys. Chem. B*, 2000, 104:3668-3675.

Wooley, K.L.; Hawker, C.J.; Fréchet J.M.; "Hyperbranched Macromolecules via a Novel Double-Stage Convergent Growth Approach", *J. Am. Chem. Soc.*, 1991, 113:4252-4261.

Yonetani, Y.; "A severe artifact in simulation of liquid water using a long cut-off length: Appearance of a strange layer structure", *Chem. Phys. Lett.*, 2005, 406:49-53.

Zeng, F.; Zimmerman, S.C.; "Dendrimers in Supramolecular Chemistry: From Molecular Recognition to Self-Assembly", *Chem. Rev.*, 1997, 97:1681-1712.

Zifferer, G.; "Shape distribution and correlation between size and shape of star-branched tetrahedral lattice chains in athermal and theta systems", *J. Chem. Phys.*, 1999, 110:4668-4677.