Molecular modelling and simulation of retroviral proteins and nanobiocomposites

Dissertation zur Erlangung des
naturwissenschaftlichen Doktorgrades
der Julius-Maximilians-Universität Würzburg

vorgelegt von
Sergey Shityakov
aus Novgorod the Great, Russland

Würzburg 2011

Eingereicht am:

Mitglieder der Promotionskommission:

Vorsitzender:

1. Gutachter: Prof. Dr. Thomas Dandekar

2. Gutachter: Prof. Dr. Axel Rethwilm

Tag des Promotionskolloquiums:

Doktorurkunde ausgehändigt am:

List of contents

List of tables

List of figures

General introduction

Molecular modelling and simulation of versatile biological systems is a very powerful toolbox in modern bioinformatics, and enables to follow and understand structure and dynamics with extreme detail on scales where motion of individual atoms can be tracked. Different biological systems such as lipid membrane bilayers and retroviral proteins play an important role in physiology and pathogenesis of living cells. The availability of high-resolution crystallographic structures, together with the development of detailed atomic models and molecular dynamics algorithms, provide a unique opportunity to refine our understanding of these systems. Although the complexity of biological structures does present a formidable challenge to theoretical studies, even with modern computational resources, it is particularly encouraging to note that many of the recent results from simulations have been consistent with the information emerging from higher resolution structural data. This relative success relies for a large part on computational strategies involving docking studies and molecular dynamics simulations.

This thesis focuses on the most commonly used methods, namely, molecular dynamics simulations, homology modelling and molecular docking which, respectively, optimize structures and simulate the natural motion of biological macromolecules with various materials and dock ligands to protein targets. The common theoretical framework based on statistical mechanics is covered briefly as well as limitations of the computational approach, for instance, the lack of quantum effects and limited timescales.

Computational techniques provide other options for understanding chemical systems, which yield information that is difficult, if not nearly impossible, to obtain in laboratory analysis. The advent of large genome sequencing reinforced the observation that structural information is needed to understand the detailed function and mechanism of biological molecules such as enzyme reactions and molecular recognition events. Molecular modelling of proteins aims primarily at establishing sequence-structure-function relationships for biological molecules using *in silico* techniques. This discipline emerged about 40 years ago (Levitt, 2001) and has made much progress in the past decade. Knowledge of the three-dimensional (3D) structures of proteins provides invaluable insights into the molecular basis of their func-

tions. Computational methods for predicting the 3D structures of proteins enjoy a high degree of interest and are the focus of many research and service development efforts. Despite an intensive effort in molecular modelling to account for the extensive thermodynamics data on proteins in pure water and solutions, some protein structures remain incompletely understood at the molecular level. Sometimes, there is no substitute for practical laboratory experience, but computer modelling methods play an important role both as an aid in interpreting experimental results and as means of explaining these results. Molecular modelling is now used not just in chemistry but in a wide range of subjects such as pharmacology, biology and biophysics. One example of the protein computer modelling approach is to provide a reasonably accurate first guess at a structure which can then be used in methods such as X-ray diffraction of powders which, unlike the X-ray diffraction of single crystals, do not provide enough information to determine the total structure from scratch.

In this thesis, my own results and relevant literatures on different important aspects of relevant molecular modelling methods are reviewed and discussed. The physiological function and chemical properties of different biological systems might be modelled with various sets of following techniques such as homology modelling, docking studies, molecular dynamics etc.

The remaining chapters of the thesis are organized as follows. In Chapter 1, detailed results and review of relevant literature on different important aspects of structural and docking analysis of HIV-1 integrase and Transportin-SR2 interaction, modelling and simulation techniques are described. Adopted experimental methodology to analyze the role of the central polypurine tract in retroviral nuclear import in context of a foamy virus vector background is described in Chapter 2. Chapter 3 is devoted to the investigation of the HIV protease inhibitor Indinavir derivatives using pharmacophoric - shape similarity scoring function. Chapter 4 focuses on the CNT interactions and molecular dynamics simulation of palmitoyloleoylphosphatidylcholine and palmitoyloleoylphosphatidylethanolamine membrane bilayers enforced by an intercalated single-wall carbon nanotube. Finally, concluding remarks from all the studies along with recommendations for future research are summarized in concluding discussion section.

Motivation for present research

Molecular modelling and simulation are powerful methods in providing important information on different biological systems to elucidate their structural and functional properties, which cannot be determined in experiment. These methods are applied to analyse the versatile biological systems: POPC/POPE lipid membrane bilayers stabilized by an intercalated single wall carbon nanotube and retroviral proteins such as HIV protease and integrase.

HIV-1 integrase has nuclear localization signals (NLS) which play a crucial role in nuclear import of viral preintegration complex (PIC). However, the detailed mechanisms of PIC formation and its nuclear import transport are still unknown. Previously it was shown that NLSs bind to the cell transport machinery e.g. proteins of nuclear pore complex such as transportins. I investigated the interaction of this viral protein HIV-1 integrase with proteins of the nuclear pore complex such as transportin-SR2. I showed the possible reasons for role of the transportin-SR2 in a nuclear import via its interaction with the HIV-1 integrase. I analyzed key domain interaction and hydrogen bond formations in transportin-SR2. These results were discussed in comparison to other retroviral species such as foamy viruses to better understand this specific and efficient retroviral trafficking route.

Experimentally, the retroviral nuclear import was analyzed by investigating the retroviral ability to infect nondividing cells. To accomplish gene transfer task successfully, retroviruses must efficiently transduce different cell cultures at different phases of cell cycle. However, promising and safe foamy viral vectors used for gene transfer therapy are unable to efficiently infect quiescent cells. This drawback was due to their inability to create a preintegration complex (PIC) for nuclear import of retroviral DNA. On the contrary, the lentiviral vectors are not dependant on cell cycle. In the course of reverse transcription the polypurine tract (PPT) is believed to be crucial for PIC formation.

In this thesis, I compared the transduction frequencies of PPT modified FV vectors with lentiviral vectors in nondividing and dividing adenocarcinomic human alveolar basal epithelial cells (A549) by using molecular cloning, transfection and transduction techniques and several other methods. In contrast to lentiviral vectors, FV vectors were not able to efficiently transduce nondividing cell in my hands. Despite these findings, which support the use

of FV vectors as a safe and efficient alternative to lentiviral vectors, there is still a big limitation in terms of foamy-based retroviral vector gene therapy in quiescent cells.

Many attempts have been made in the recent years to search for the potential new drugs to treat HIV infection. These molecules can be retrieved from chemical libraries or can be designed on a computer screen and then synthesized in a laboratory. Most notably, one could use the computerized structure as a reference to determine the types of molecules that might block the enzyme. Such structure-based drug design strategies have the potential to save off years and millions of dollars compared to a more traditional trial-and-error drug development process.

After the crystal structure of the HIV-encoded protease enzyme had been elucidated, computer-aided drug design played a pivotal role in the development of new compounds that inhibit this enzyme which is responsible for HIV maturation and infectivity. Promising representatives of these compounds have recently found their way to patients. Protease inhibitors show a powerful sustained suppression of HIV-1 replication, especially when used in combination therapy regimens. However, these drugs are becoming less effective to more resistant HIV strains due to multiple mutations in the retroviral proteases.

In computational drug design I used molecular modelling methods such as lead expansion algorithm (Tripos®) to create a virtual library of compounds with different binding affinities to protease binding site. In addition, I heavily applied computer assisted combinatorial chemistry approaches to design and optimize virtual libraries of protease inhibitors and performed *in silico* screening and pharmacophore-similarity scoring of these drug candidates. Further computational analyses revealed one unique compound with different protease binding ability from the initial hit and its role for possible new class of protease inhibitors is discussed in the appropriate chapter.

A number of atomistic models were developed to elucidate the nanotube behaviour in lipid bilayers. However, none of them provided useful information for CNT effect upon the lipid membrane bilayer for implementing all-atom models that will allow us to calculate the deviations of lipid molecules from CNT with atomistic precision. Unfortunately, the direct experimental investigation of nanotube behaviour in lipid bilayer remains quite a tricky problem opening the door before the molecular simulation techniques. In this regard, more detailed multi-scale simulations are needed to clearly understand the stabilization characteristics of CNTs in hydrophobic environment.

The phenomenon of an intercalated single-wall carbon nanotube in the center of lipid membrane was extensively studied and analyzed. The root mean square deviation and root

mean square fluctuation functions were calculated in order to measure stability of lipid membranes.

The results indicated that an intercalated carbon nanotube restrains the conformational freedom of adjacent lipids and hence has an impact on the membrane stabilization dynamics. On the other hand, different lipid membranes may have dissimilarities due to the differing abilities to create a bridge formation between the adherent lipid molecules. The results derived from this thesis may be of importance in developing stable nanobiocomposites for construction of novel biomaterials and delivery of various biomolecules in the field of materials science.

Chapter 1

1.1 Overview

HIV-1 integrase has NLS (nuclear localization signals) which plays an important role in internuclear transport of viral PIC (preintegration complex). The exact mechanisms of PIC formation and its internuclear transport are not known. It was shown that NLSs bind to the cell transport machinery e.g. proteins of nuclear pore complex such as transportins. I investigated the interaction of this viral protein with proteins of the nuclear pore complex (transportin-SR2). I showed reasons why transportin-SR2 is the nuclear import protein for HIV-1 integrase and not transportin-SR1: (i) 3D alignments identify differences between transportin-SR1 and transportin-SR2. (ii) Rigid protein-protein docking showed key domain interactions and hydrogen bonds available to transportin-SR2. (iii) Flexible receptor-ligand docking was performed to reveal crucial amino acid residues involved in this hydrogen bond formation. These results are discussed to better understand this specific and efficient retroviral transport route comparing the interactions of related retroviruses (SIV, HIV-2, PFV etc.) with their cognate transport proteins, NLS sequences and kinase binding motifs.

1.2 The problem to solve

In contrast to other viruses, retroviruses generate complex and structured pre-integration complexes (PICs) to transport reverse transcribed RNA genomes into the host nucleus. In addition to cellular proteins, HIV-1 PIC consists of viral proteins such as reverse transcriptase (RT), integrase (IN), matrix protein (MP), auxiliary protein Vpr and viral DNA with a flap region. The PIC has the ability to bind proteins of the nuclear pore complex (NPC) to enter the nucleus and integrate into the host genome (Bushman *et al.*, 1990; Engelman *et al.*, 1991). To penetrate through nuclear membrane PIC proteins must bear nuclear localization signals (NLS) for recognition by the proteins of the nuclear pore complex.

Numerous studies have analyzed NLS and PIC interactions, but the exact mechanisms of PIC formation and its nuclear import are unknown. It was shown previously (Christ *et al.*,

2008) that the HIV-1 IN NLS binds to the cell transport machinery factors, e.g. proteins of the nuclear pore complex such as transportins. Transportins (TR-SRs) are serine/arginine-rich proteins. They are members of the kariopherin family and play an important role in nuclear import, RNA-splicing regulation, and signalling transduction (Lai *et al.*, 2003). TR-SRs are superhelical S-like molecules consisting of helically stacked HEAT repeats. These 37-47 amino acid long modules are randomly repeated (Chook and Blobel, 1999). TR-SRs have two domains attached to each other by an acidic loop. Domain one is N-terminal Ran-GDP binding (Ran-GDP BD, 1-305aa) and the other is C-terminal cargo binding (CBD, 381-891aa).

The Ran-GTPase is a nucleocytoplasmic shuttle protein and plays a key role in promotion of nuclear transport in both directions (Izaurralde *et al.*, 1997). This simplified whole nuclear transport process has four stages: i) binding of the cargo (HIV-1 IN) in the cytosol by the receptor (TR-SR), ii) movement of the complex through the NPC into the nucleus, iii) Ran-GTP binding to complex that in turn promotes release of the cargo, and iv) export of the complex to the cytosol, with hydrolysis and release of RanGDP in the cytosol (Mattaj and Englmeier, 1998).

There are different homologous variants of transportin: transportin-SR1 (TR-SR1) and transportin-SR2 (TR-SR2). TR-SR1 was first identified as a nuclear transport receptor for heterogeneous nuclear ribonucleoprotein A1 (hnRNP A1) (Siomi and Dreyfuss, 1995; Pollard *et al.*, 1996; Iijima *et al.*, 2006). TR-SR2 is subdivided into isoform A and isoform B. TR-SRs are referred to be redundant factors for the nuclear import of peptides such as M9, HNS as well as proteins (e.g. HuR, hnRNP A1, D, JKTBP, TAP) and DNA such as viral PICs (Rebane *et al.*, 2004; Christ et al., 2008; Imasaki *et al.*, 2007).

HIV-1 integrase (HIV-1 IN) is a 288 amino acid protein which consists of a His_2Cys_2 zinc-finger motif at its N-terminus, a catalytic core domain (CCD) in the center of the molecule, and a DNA-binding domain at its C-terminus **(Figure 1.1)**. There is a NLS in the center of the catalytic core domain (Bouyac-Bertoia *et al.*, 2001), which probably plays a role in IN recognition by TR-SR. HIV-1 IN CCD consists of five β-sheets with six α-helices. The β-sheets are hidden inside of the CCD. They are creating a kind of 'core', whereas the α-helices are opposed outside and are forming a 'shell-like' structure. These α-helices contain the NLS as well (Bouyac-Bertoia *et al.*, 2001).

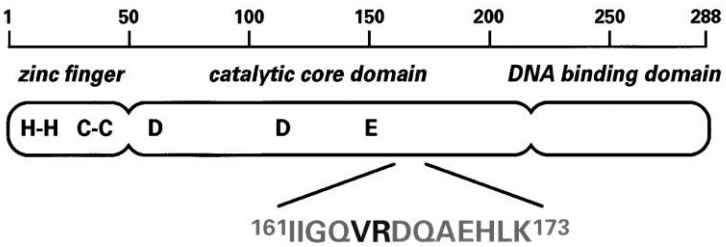


Figure 1.1: The three-domain structure of HIV-1 IN. The sequence of the 161–173 region of the HIV-1 virus is shown and the two residues essential for NLS function are indicated in black (figure modified after Bouyac-Bertoia et al., 2001).

In vitro studies indicated that HIV-1 IN forms dimers. It has been proposed that these dimers assemble into the fully functional tetramers to position two active sites within 15 Å (Dyda *et al.*, 1994), but the exact number of monomers in the whole functional enzyme still needs to be determined.

Nucleocytoplasmic traffic is a dynamic process and implies different receptors, ligands and complex pathways with many unknowns. Recently, the mechanism of HIV-1 nuclear import has been studied in terms of HIV-1 IN interaction with LEDGF/p75 (lens epithelial-derived growth factor), which is a cellular co-factor for HIV-1 replication. Its role is the targeting of viral IN to the host chromosome (Emiliani *et al.*, 2005). Another cellular component, TR-SR2 probably is important for transport of the viral PIC through the NPC (Christ *et al.*, 2008).

Here, I analyze the following questions: Which are the TR-SR domains and motifs that will possibly interact with HIV-1 IN? What is the structural basis for this interaction? I examine also the generality of this retroviral nuclear import mechanism in terms of uniqueness regarding to NLSs in other retroviruses, regulatory phosphorylation and implications for anti-retroviral therapy and retroviral evolution.

1.3 Computational methods

1.3.1 Structural analysis software

A TR-SR2 model was made using the ESyPred3D automated homology modelling server (Lambert *et al.*, 2002). The build-in neural network created multiple alignments to enhance performance of the integrated modelling package (MODELLER).

The ESyPred3D automatic program implements the four steps for homology modelling approach: 1) databank searching to identify the structural homology; 2) target-template alignment; 3) homology model building and optimization; 4) model evaluation. To find the homolog to the target sequence, PSI-BLAST was implemented in search. The ESyPred3D cutoff value was 0.0001.

The crystal structure of the human TR-SR1 molecule was used as a template (pdb code: 2Z5J). The three dimensional structure of the catalytic core domain of the HIV-1 IN (protein data bank code: 1BI4) was retrieved from Protein Data Bank and refined using the NOC 3.0 software.

The VADAR server, which is an improved version of the PROCHECK software, was used for stereochemical validation of the homology model (Willard *et al*., 2003). During the determination and refinement of a structure, regions in error were identified by stereochemical considerations. Standard checks include the identification of bad contacts, the careful inspection of the structure on a graphics terminal and the use of a Ramachandran plot to see which residues lie in the 'allowed' or 'disallowed' regions.

Amino acid chemical properties and motif predictions were calculated by the NOC 3.01 software. The molecular modifications (TR-SR2 H8-loop, HIV NLS) were created by the PyMol Protein Editor and the BALLView 1.2 software (Moll *et al*., 2005).

The Molprobity and the FireDock servers were used for structural and docking refinements. The MolProbity session starts with the user uploading a coordinate file of their own or fetching one from the PDB databases (Berman *et al.*, 2000) in protein data bank format or in the macromolecular crystallographic information file format. After checking the thumbnail image and listed characteristics of the pbd file, hydrogen atoms were added and optimized, with automated correction of Asn/Gln/His 180° flips (Richardson *et al.*, 1999).

The FireDock refinement process consisted of two main steps: (1) rearrangement of the interface side chains and (2) adjustment of the relative orientation of the molecules. The FireDock binding score (binding free energy) was calculated as the change in the free energy of the system, which occurs upon complex formation,

$$\Delta G = G^C - (G^R + G^L)$$

where G^C is the free energy of the receptor-ligand complex. G^R and G^L are the free energies of the receptor apo structure and ligand, respectively. The binding score for the candidates ranking is an approximation of the binding-free energy function.

Dijkstra's algorithm of the CAVER in the PyMol was implemented in searching process and started from source node (starting point), which located in the center of the molecule. The algorithm developed automatically finds the easiest path from the starting point, typically located inside the molecule, to the exterior of the molecule. The identified path resembles a tunnel that connects protein clefts, pockets or cavities with the surrounding bulk solvent. The tunnel characteristics, e.g. length, mean radius and gorge radius are determined and can be further analyzed **(Figure 1.2)**. The tunnel gorge radius r_{gorge} is one of the most important tunnel characteristics because the tunnel gorge can form a bottleneck for substrate access or product release to and from the active site of a protein. The maximal error "max" of r_{gorge} estimation is expressed by the equation:

$$\varepsilon_{max}\left(r_{gorge}\right) = \frac{\sqrt{3}}{2}d$$

where d is equal to the length of the grid cell edge. The probability of ε_{max} realization is equal to zero and this error is overestimated, therefore the mean error should be defined. The mean error of r_{gorge} determination is equal to

$$\langle\varepsilon_{max}\left(r_{gorge}\right)\rangle = \frac{1}{V}\iiint_{d^3} rdV \cong 0.48d$$

and its variance and deviation equal to

$$var(\varepsilon) = \langle\varepsilon_{max}^2\left(r_{gorge}\right)\rangle - \langle\varepsilon_{max}^2\left(r_{gorge}\right)\rangle^2$$

$$var(\varepsilon) \cong 0.01d^2, \sigma\varepsilon \cong 0.14d$$

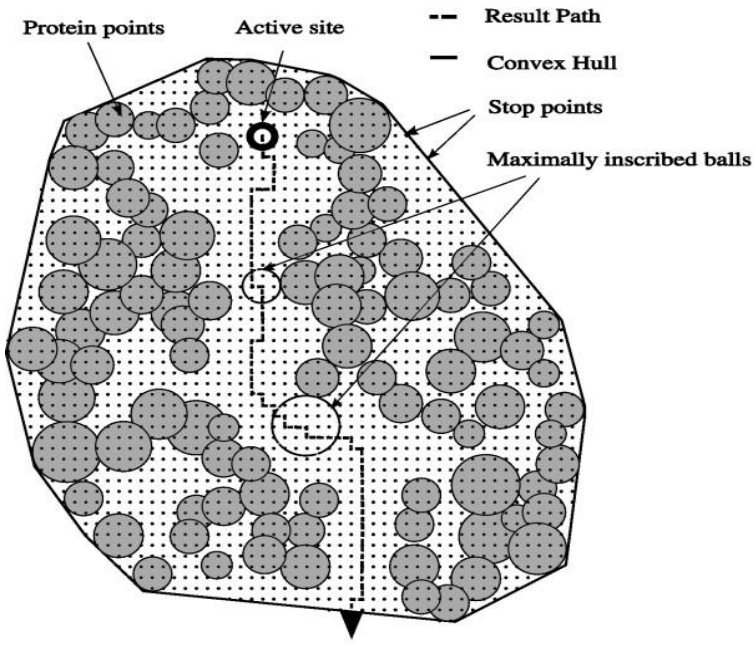


Figure 1.2: Sketch of the computational algorithm implemented in the CAVER. The black bold circle represents the starting point. The protein is visualized by gray circles with Van der Walls atom radii mapped on a discrete grid (black dots). The solid line represents the boundary between the protein (convex hull) interior and its surroundings. Empty circles represent the maximally inscribed balls on the probable route (dashed line).

1.3.2 Sequence analysis software

Alignments and superimpositions were made by means of the extensible molecular modelling system UCSF Chimera, alpha version 1.3.

Molecular graphics images were also produced using the UCSF Chimera package from the resource for biocomputing, visualization and informatics at the University of California, San Francisco (Petterson *et al.*, 2004). The molecular graphics program UCSF Chimera includes a suite of tools for interactive analyses of sequences and structures. Structures automatically associate with sequences in imported alignments, allowing many kinds of crosstalk.

Advantages of using Chimera for sequence/structure research include a rich set of features co-existing in a single program, certain unique methods, facile integration of user data, access to the broader visualization and analysis capabilities of Chimera, program extensibility and detailed documentation.

1.3.3 Docking programs

In this study, I used the PatchDock (molecular docking algorithm based on shape complementarity principles) web server v.1.2.0 together with the AUTODOCK 4.0 suite to perform rigid protein-protein and flexible receptor-ligand dockings, respectively.

PatchDock (Duhovny *et al.*, 2002) is a geometry-based molecular docking algorithm. It is aimed at finding docking transformations that yield good molecular shape complementarity. Such transformations, when applied, induce both wide interface areas and small amounts of steric clashes. A wide interface is ensured to include several matched local features of the docked molecules that have complementary characteristics. The PatchDock algorithm divides the Connolly dot surface representation (Connolly, 1983) of the molecules into concave, convex and flat patches. Then, complementary patches are matched in order to generate candidate transformations. Each candidate transformation is further evaluated by a scoring function that considers both geometric fit and atomic desolvation energy.

AUTODOCK 4.0 is a package of three main C++ programs: (1) AutoTors implies the input of ligand parameters; the (2) AutoGrid is used to calculate a three-dimensional grid of interaction energy based on macromolecular coordinates; (3) AutoDock creates the molecular docking (Goodsell and Olson, 1990) and recruits the AMBER (Assisted Model Building with Energy Refinement) force field. The program AutoDock was developed to provide an automated procedure for predicting the interaction of ligands with biomacromolecular targets. The original search algorithm is the Metropolis method, also known as Monte Carlo simulated annealing. With the protein static throughout the simulation, the substrate molecule performs a random walk in the space around the protein. At each step in the simulation, a small random displacement is applied to each of the degrees of freedom of the substrate: translation of its center of gravity; orientation; and rotation around each of its flexible internal dihedral angles. This new energy is compared to the energy of the preceding step. If the new energy is lower, the new configuration is immediately accepted. In case of high energy profile, the configuration is accepted or rejected based upon a probability expression dependent on defined temperature, *T*. The probability of acceptance is given by:

$$P(\Delta E) = e^{\left(-\frac{\Delta E}{K_B T}\right)}$$

where ΔE is the difference in energy from the previous step, and K_B is the Boltzmann constant. At high enough temperatures, almost all steps are accepted. At lower temperatures, fewer high energy structures are accepted. The simulation proceeds as a series of cycles, each at a specified temperature. Each cycle contains a large number of individual steps, accepting or rejecting the steps based upon the current temperature. After a specified number of acceptances or rejections, the next cycle begins with a temperature lowered by a specified schedule such as:

$$T_i = gT_{i-1}$$

where T_i is the temperature at cycle i, and g is a constant between 0 and 1.

Simulated annealing allows an efficient exploration of the complex configurational space with multiple minima that is typical of a docking problem. The separation of the calculation of the molecular affinity grids from the docking simulation provides modularity to the procedure, allowing the exploration of a range of representations of molecular interactions, from constant dielectrics to finite difference methods and from standard potential functions to distributions based on observed binding sites.

1.4 Results and discussion

1.4.1 Highly similar sequences of TR-SR2 and TR-SR1 have different 3D binding properties and domain structures

Sequence alignment: I aligned TR-SR1 (835 aa) and TR-SR2 (887 aa) sequences. The molecules have 91.4% similarity and 82.0% identity over the alignment length, 87.1% identity over the aligned residues, and 87.1% identity over shorter sequence. The main and all chain aligned root mean square deviations (RMSDs) were 7.86 and 8.26 Å, respectively **(Figure 1.3)**.

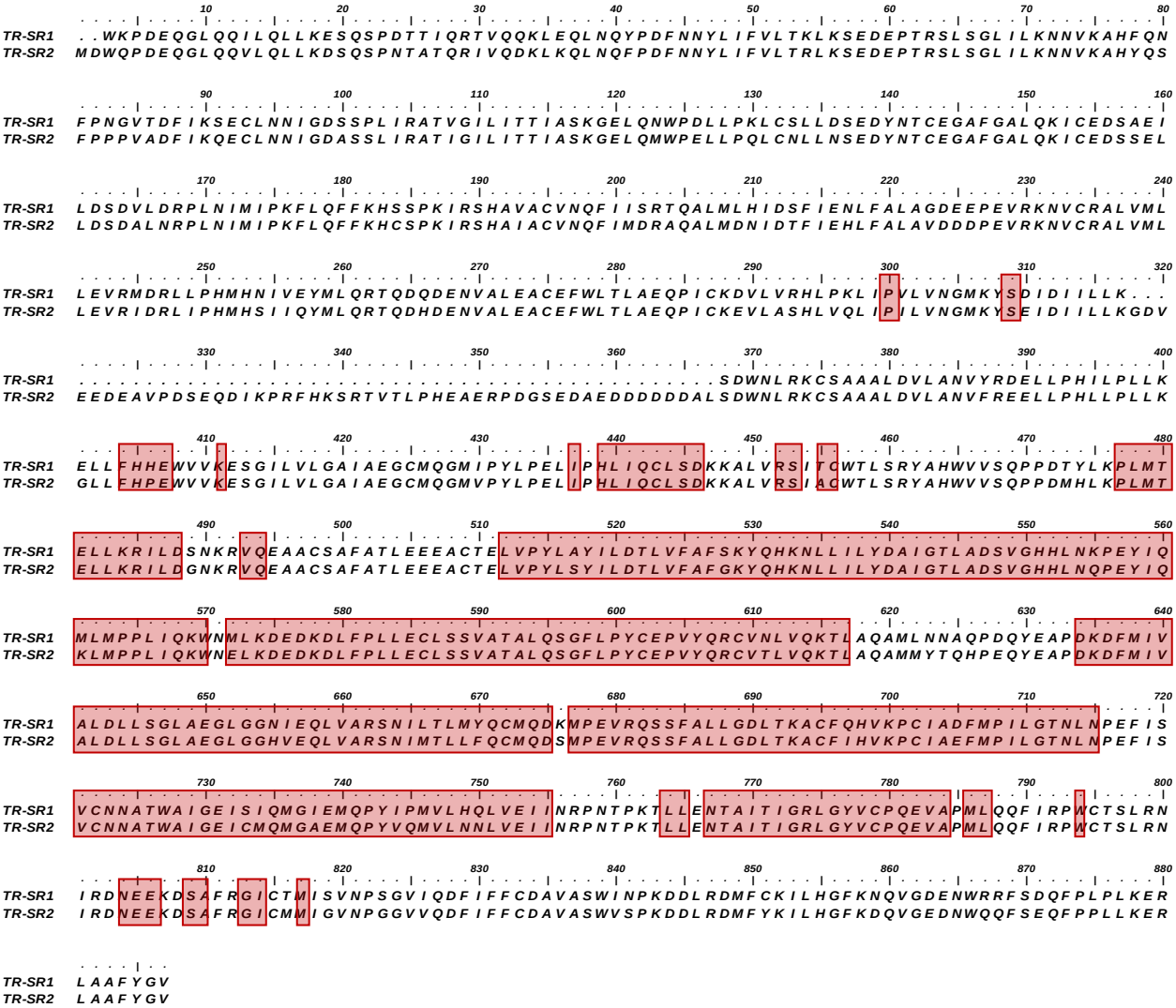


Figure 1.3: Sequence alignments and superimposition of TR-SR1 and TR-SR2 molecules. The root mean square (RMS) value calculated with the UCSF Chimera v.1.3 (see section Materials and Methods) is 2.0 Å between the two structures: crystal TR-SR1 and TR-SR2 (homology model). The superimposed regions are marked in red (see 1.3 section for details).

Sequence analysis of different domains showed 83.8% of total similarity between TR-SR1 and TR-SR2 Ran-GDP BDs (1-303 aa) and 88.4% between CBDs (379-887 aa), respectively.

The obtained TR-SR2 homology model was validated stereochemically. Altogether 785 (88%) of all residues were in favoured and allowed regions.

Structure alignment: To compare folding structures of the two transportin molecules, I performed the three-dimensional superimposition **(Figure 1.4 A)**. No similarity was found by superimposition of the TR-SR1 and TR2-SR2 Ran-GDP BDs (1-303 aa; cutoff: RMSD of 2 Å). However, the TR-SR1 and TR-SR2 cargo binding domain comparison (379-887 aa) revealed 57.7% of similarity.

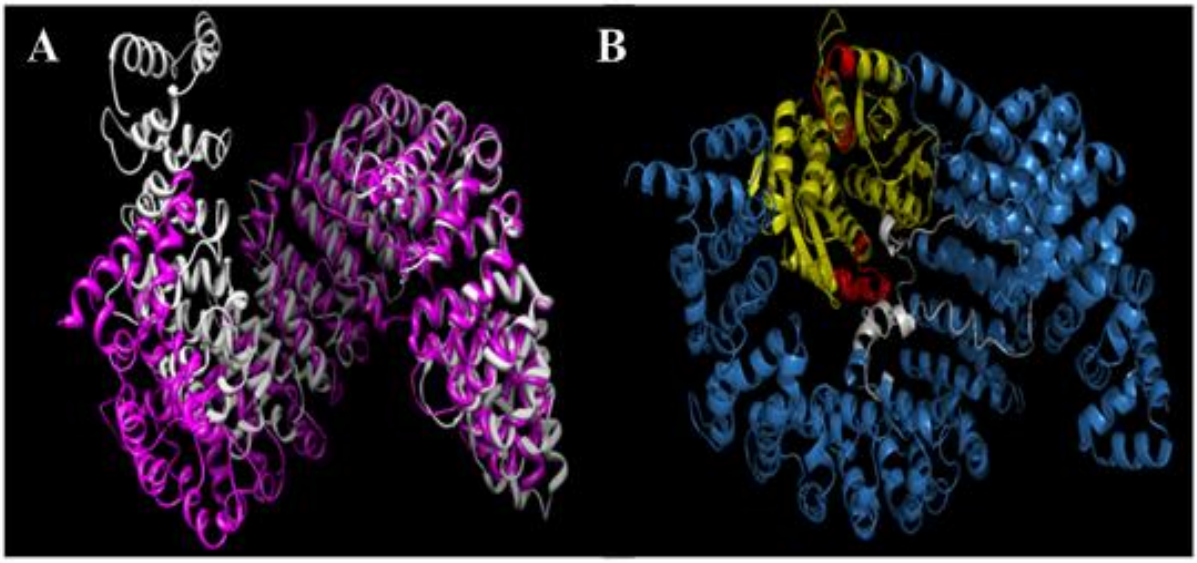


Figure 1.4: (A) TR-SR1 and TR-SR2 superimposition, coloured in gray and blue respectively. Superimposition was performed on the basis of Needleman-Wunsch-like alignment algorithm. Scores for aligned characters were specified by similarity substitution matrix (BLOSUM-62), which is widely used to score alignments between evolutionary divergent protein sequences (Needleman and Wunsch, 1970; Henikoff et al., 1992). Sequences were iterated by pruning long atom pairs until no pair exceeded 2.0 Å RMSD. (B) The rigid docking of HIV-1 IN CCD and TR-SR2. Protein domains painted in different colours: HIV-1 IN CCD (162 aa) is green; NLS (13 aa) is red; Ran-GDP BD (1-303 aa) is grey; Acidic loop (304-378 aa) is yellow; CBD (379-887 aa) is blue, RMSD=1.6 Å. The pictures were generated using the PyMol software.

One possibility for these structure dissimilarities might be in the dihedral angles of the acidic loop, which is tilting the domains and hence increases structural RMSD levels. To investigate this, the dihedral angles of the acidic loop central atom for TR-SR1 and TR-SR2 were measured. Φ and ψ angles were comprised of -115°, -44° for TR-SR1 and -64°, -40° for TR-SR2 respectively. To investigate further the discrepancy between high sequence identity and different 3D domain structures (high structural RMSD) of Ran-GDP and cargo-binding domains, I performed cutoff 3D alignment (cutoff: RMSD of 2 Å) of the domains on their own and found 62.2% identity for CBDs and only 14.8% identity for Ran-GDP binding domains. These data support the idea that 3D differences might be due to inherent structural domain organization differences rather than just the tilting of the molecules. These structural differences despite high sequence similarity help to explain results from yeast two-hybrid experiments and pull-down assays of mammalian cell extracts where the HIV-1 IN binds to the TR-SR2, whereas it does not interact with TR-SR1 (Christ *et al*., 2008).

1.4.2 Where does HIV-1 integrase bind to TR-SR2?

Imasaki and his co-authors allowed one to identify the Ran-GTP and the two cargo binding sites in TR-SR1. With nuclear import governed by such a mechanism, the role of the Ran-GDP binding domain in HIV-1 IN NLS recognition is not clear, and why one may dock these two structures? However, a possible reason for this is a region of TR-SR1, the H8-loop, that authors indicate, might play a role in both Ran-GTP binding and in NLS release through what they term a "pre-unloaded" state of the complex. According to this knowledge, I docked the HIV-1 integrase catalytic core domain (HIV-1 IN CCD) to the H8-loop of TR-SR2 to examine the possible flexibility of docked molecules by the PatchDock server. The PatchDock server provides an opportunity to specify the interactor chain IDs of the receptor (macromolecule) and the ligand (Schneidman-Duhovny *et al.*, 2005). I specified the receptor and the ligand interaction sites: 161-173 aa (NLS) and 312-374 aa (H8-loop). At least two of the listed HIV-1 IN NLS residues should be in contact with any residues of the TR-SR2 molecule.

The PatchDock Server had generated 1000 solutions. The input parameters were 0.8 for coarse refinement and 0.85 for a final refinement of the best candidates. The best candidates were refined with the FireDock module and sorted according to their energetic parameters. The table was sorted by global energy values, such as the global energy, the attractive and repulsive van der Walls forces, the atomic contact forces and the hydrogen bond energy. Refined complex structures were generated for 7 top-rank conformations for further analysis

(see Table 1.1). According to refinement parameters, the minimum of global energy (-0.50 kcal/mol) was found in the first docking conformation **(Figure 1.4 B).**

Table 1.1 Protein-protein docking (kcal/mol)

Rank	Solution Number	Global Energy	Attractive VdW[1]	Repulsive VdW	ACE[2]	HBE[3]
1	6	-0.50	-20.21	15.78	3.71	0.00
2	3	-0.37	-39.79	29.36	4.69	-5.05
3	2	2.06	-1.72	0.00	0.12	-0.80
4	5	5.19	-22.87	9.68	-4.69	0.00
5	4	9.72	-25.44	19.09	6.44	-1.79
6	7	13.03	-6.27	0.14	4.51	0.00
7	1	14.98	-3.89	1.69	-1.13	0.00

[1]*-Van der Waals Energy*
[2]*-Atomic Contact Energy*
[3]*-Hydrogen Bond Energy*

1.4.3 Hydrogen bonds involved in the binding between HIV-1 IN NLS and the H8-loop of TR-SR2 Ran-GDP binding domain

Based on the rigid docking, NLS and H8-loop were chosen as candidates for flexible binding and were docked with each other. However, there are different NLS motifs in the sequences considered (Gallay *et al*., 1997). My study focuses on the conserved HIV-1 IN NLS (161-173 aa). Implicating of this motif in the PIC nuclear import was shown to be essential in retroviral replication (Bouyac-Bertoia *et al.*, 2001).

To determine amino acids important for hydrogen bond formation, I performed flexible receptor-ligand docking using the AUTODOCK 4.0 software and its Monte Carlo method. This method creates random moves to the system and then can accept or deny the move according to Bolzmann probability (Berg, 2004). Flexible protein-protein docking implements a tremendous number of rotational degrees of freedom for each amino acid. Therefore, two conflicting requirements must be well balanced in the automated docking process: the desire for robustness of physically compatible models and feasible computation (Goodsell *et al.*, 1996). Follow this rule I used TR-SR2 H8-loop as a 'receptor molecule' which is rigid and HIV-1 IN NLS as a 'ligand' which is flexible. The H8-loop and NLS were modelled as the alpha-helix structures. Their alpha-helical dihedral angles were measured on the Ramachandran plot and changed to the appropriate ones, represented in the 'native' protein structures. The H8-loop and NLS were refined by Molprobity refinement server. H8-loop was solubilized in water box by adding 127 molecules of water. The ADT tool set provides two methodologies for charge calculation: Kollman and Gasteiger charges. Appropriate

charge energies were added: total Kollman and Gasteiger charges were -14.0, -16.0005 for H8-loop and 2.002, -0.0009 for NLS respectively. Charge field was then set to Kollman. I used only 29 out of 71 total rotational bonds in HIV-1 IN NLS. One requisite for correct docking was that the amide bonds should not be rotational. I also analyzed other peptide backbone bonds non-rotational to drop the number of possible rotating bonds to 29 (maximum is 32).

The construction of "affinity grids" for each type of atom in the substrate/"donor" is particularly crucial to a successful dock, for this reason I analyzed buried volume of the H8-loop using the PyMol CAVER module, to identify certain atomic positions of hidden binding moieties.

The H8-loop binding center **(Figure 1.5 A)** was detected with coordinates (x, y, z): -0.720, -3.060, -2.280 (center of the receptor: -0.848, -3.087, -1.777; center of the ligand: 0.972, 0.937, -0.957). Number of points in three dimensions was adjusted to 54, 40 and 60 with 0.375 Å spacing and 137555 current total grid points per map. For preparing the docking parameter file, I used default settings (genetic algorithm parameters: population size = 150, number of energy evaluations = 2500000, rate of gene mutation = 0.02, rate of crossover = 0.8, maximum number of generations = 27000, number of GA runs = 10, number of active torsions in ligand = 31, number of torsional degrees of freedom = 55, coefficient of torsional degrees of freedom = 0.274, initial dihedrals were randomly specified, elitism value was set to 1). By default, clustering of docked results was done at 0.5 Å RMSD (Table 1.2). The number of conformations by default was 10.

Table 1.2 Docking clustering (kcal/mol)

Cluster Rank	Lowest Docked Energy	Run	Mean Docked Energy	Number of clusters
1	+4.84	2, 9	+5.97	2
2	+5.01	7	+5.01	1
3	+5.44	10	+5.44	1
4	+5.59	3	+5.59	1
5	+5.60	8, 6	+7.24	2
6	+6.39	5, 9	+7.24	2
7	+7.36	4	+7.36	1

Table 1.3 Docking conformations (kcal/mol)

Conformation Run	Free energy of binding	vdW+Hbond +dissolve energy	Electrostatic energy	Final Total internal energy	Final Intermolecular energy
1	+8.09	-3.43	-0.88	-3.51	-4.31
2	+4.84	-4.33	-2.41	-4.33	-6.74
3	+5.59	-5.74	-1.55	-3.02	-7.30
4	+7.36	-3.13	-1.19	-4.32	-4.32
5	+6.39	-3.64	-2.76	-3.11	-6.40
6	+8.87	-4.32	-1.05	-1.66	-5.38
7	+5.01	-5.62	-2.01	-3.27	-7.63
8	+5.60	-4.11	-2.66	-3.54	-6.77
9	+7.09	-3.77	-0.20	-4.85	-3.96
10	+5.44	-6.27	-1.53	-2.66	-7.81

Table 1.3 shows various HIV-1 IN NLS ('ligand') conformations as a result of differences in the docked energy. The scoring function (Poisson-Bolzmann method) was applied to diminish the number of false positive conformations. The second conformation had reached a global minimum in the docked energy (+4.84kcal/Mol) at 2.0 Å RMSD. Estimated free energy of binding was calculated as a sum of the final intermolecular energy, the final total internal energy, the torsional free energy and the unbound system energy. The torsional free energy (+15.09 kcal/mol) and the unbond system's energy (-0.82 kcal/mol) were constant parameters for all of 10 conformations.

Table 1.4 Close contact residues that could be involved in the H-bond formation (This data obtained from AUTODOCK 4.0 calculations)

donor 'ligand' ⟹ acceptor 'receptor' residues in close contact	donor 'receptor' ⟹ acceptor 'ligand' residues in close contact
Ile 161 ⟹ Asp 332, Lys 334,	Lys 334 ⟹ Ile 161
Ile 333, Asp 328	Arg 336 ⟹ Gln 168, Val 165
Gln 168 ⟹ Arg 336	Asp 361 ⟹ Arg 166
His 171 ⟹ Ser 355	Asp 328 ⟹ Ile 161
Val 165 ⟹ Arg 336	Asp 332 ⟹ Ile 161
Arg 166 ⟹ Asp 361	Ser 355 ⟹ His 171
	Ile 333 ⟹ Ile 161

The conformation with minimal free energy of binding was chosen as approximated structure for further evaluation (Table 1.3). Hydrogen bonds were found between Ile 161 (donor), Asp 328 (acceptor); Arg 336 (donor), Gln 168 (acceptor); Arg 166 (donor), Asp 361 (acceptor) **(Figure 1.5 B)**. Moreover, there are additional residues that could form hydrogen bonds. Those amino acids, which are in close contact with each other (Table 1.4), should be

considered as molecules with possible hydrogen bond formations. Both molecules could have donor and acceptor residues.

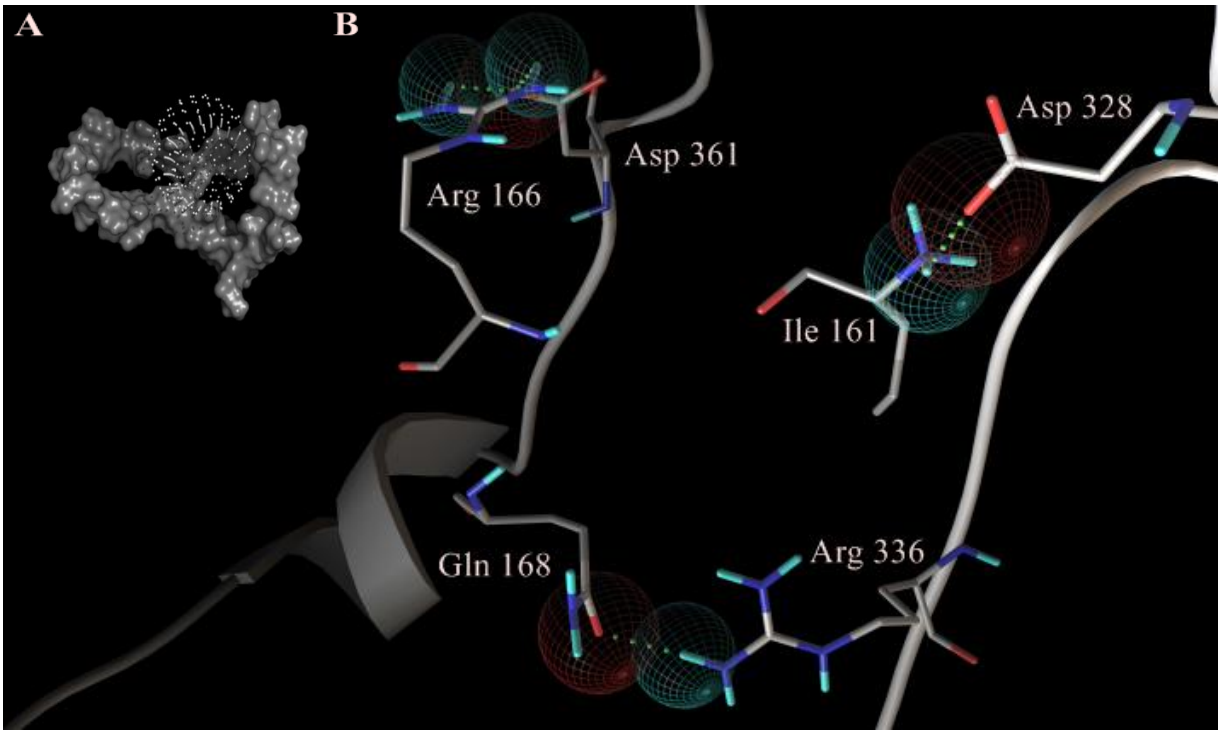


Figure 1.5: (A) H8-loop binding center detected by CAVER module. (B) The second conformation of HIV-1 IN NLS and TR-SR2 H8-loop flexible docking. Labelled amino acids are forming H-bonds between two molecules (see text for details). Molecules are coloured according to their atom composition. Wire frame spheres are displayed on the atoms with the hydrogen bonds. The pictures were generated using the PyMol software.

1.4.4 Role of highly accessible and hydrophilic amino-acids in this protein interaction

In the next step, I analyzed amino acid composition of the HIV IN NLS and the TR-SR2 H8-loop (**Table 1.5**) by calculating amino acid accessibility residues and measuring hydrophophilicity. Hydrophilic amino acids are usually exposed on the protein surface. Canonical amino acids involved in binding are basic, polar, positively charged and very hydrophilic (Creighton, 1993). As shown in **Table 1.5**, H8-loop Asp 328 is a highly accessible residue (80.3% accessibility score) that interacts with NLS Ile 161 (scores of 92.2%); Arg 336 (86.4%) interacts with Gln 168 (66.9%) and Asp 361 (64.2%) with Arg 166 (38.5%). Furthermore, all the molecules excluding isoleucine are very hydrophilic (hydrophobicity index = -4.5 and -3.5 respectively) **(Figure 1.6)**.

Table 1.5 Residue's accessibility surface area and hydrophobicity index of docked molecules (see text for details).

H8-loop	Area	Acc. (%)	HI[1]	IN NLS	Area	Acc. (%)	HI[1]
Asp 328	130.255	80.3	-3.5	Ile 161	157.887	92.2	4.5
Asp 332	106.409	65.6	-3.5	Val 165	71.374	48.0	4.2
Ile 333	111.333	65.0	4.5	Arg 166	93.843	38.5	-4.5
Lys 334	115.095	54.4	-3.9	Gln 168	125.090	66.9	-3.5
Arg 336	210.727	86.4	-4.5	His 171	127.643	66.6	-3.2
Ser 355	74.924	100.0	-0.8				
Asp 361	104.115	64.2	-3.5				

[1]*-hydrophobicity index*

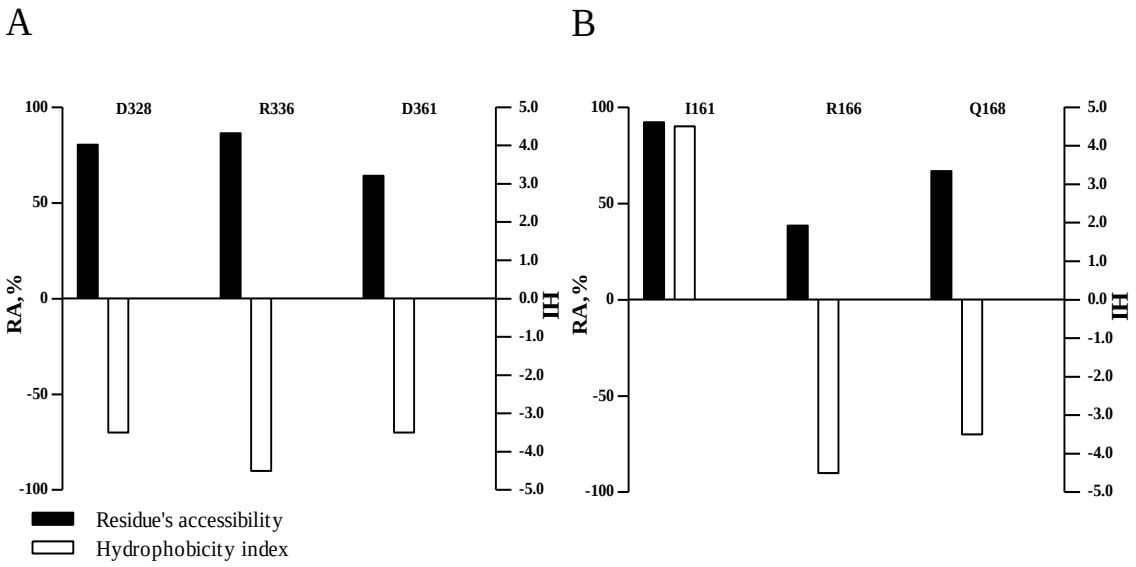


Figure 1.6: The H8-loop of TR-SR2 (A) and the HIV-1 IN NLS amino acids (B) are represented. There is a correlation between residue's accessibility and hydrophobicity of the amino acids with tendency to H-bond formations. Highly accessible amino acids have very low hydrophobicity index.

Accessibility surface area (ASA) calculation parameters in the experiment were 493 atoms, 7093.7 Å^2 overall ASA for H8-loop and 105 atoms, 1556.5 Å^2 overall ASA for NLS with probe radius = 1.40 Å.

1.4.5 Is this a general type of viral transport interaction?

I focused on the central (Bouyac-Bertoia *et al.*, 2001) HIV-1 NLS located in the CCD but there are further NLSs of the basic bipartite type (aa residues 186-188 and 211-219) and maybe unknown NLSs (Gallay *et al.*, 1997). Here I could only speculate on their possible role

in the docking of the TR-SR2. This hypothesis can be probed by the mutation analysis and further binding experiments.

The other retroviruses such as prototype foamy virus (PFV), have non-classical NLSs (basic aa signals) similar to HIV but probably more disperse. There is a region at the PFV IN C-terminal domain, which contains basic amino acids (306-329 aa). This region was recently identified as a NLS (An *et al*., 2008), and probably this signal is similar to HIV-1 NLS located in the CCD. Taken all the HIV and the PFV NLS dissimilarities into account, it is complex to compare NLS dockings of both viruses. Moreover, PFV also contains many NLSs in different parts of IN molecule (An *et al*., 2008). It is also well known that PFV replication is cell-cycle dependent and foamy viral infection of non-dividing cells is not quite clear (Bieniasz *et al*., 1995). This fact arises too many questions upon the rational of the NLSs in this type of retroviruses.

PFV docking simulation is a very hard task because of the absence of the PFV IN crystallographic structure and as an *ab initio* model is less precise then a homology model. Experimental or further information on possible PFV IN and TR-SR2 interaction is limited. Albeit retroviral integrases share a common three-domain organization, the difficulties in the HIV and PFV integrase comparison will be presumably due to the different viral domain interaction (HIV-1 CCD vs. PFV C-terminal domain) with the TR-SR2.

Regulation of this type of viral nuclear transport and therapeutic applications: Phosphorylation is important in the regulation of different cellular processes. Some authors estimate phosphorylation as a main mechanism controlling nuclear import of particular proteins such as NFkappaB, c-rel etc (Jans and Huebner, 1996). I examined here kinase binding motifs (KBM) and their involvement in viral nuclear transport regulation. Are there KBMs inside of other retroviral NLSs? What are KBM differences in different retroviruses? Are they important in nuclear import regulation? To answer these questions, I modelled the NLSs of different viruses and screened them for possible kinase binding motifs.

As shown in **Table 1.6**, most of lentiviruses (HIV, SIV) have similar IN NLSs and, hence, they have sites for phosphorylation by tyrosine kinase (pattern: [RK]-x(2,3)-[DE]-x(2,3)-Y). Interestingly, this rule does not apply for HIV-2 (cl. A), SIV/smm and SIV/syk, and I did not find any kinase binding motifs in those viruses. This phenomenon could be explained by the mutations in the intrinsic motifs or different NLSs are probably bearing functional KBMs.

Table 1.6 Different viral Nuclear Localization Signals (NLS) contain different Kinase Binding Motifs (KBMs are coloured in red).

Viral Variant	NLS	KBM
HIV-1 (YU-2)	LKKIIGQVRDQAEHLKTAV	Tyrosine kinase binding motif
HIV-1 (cl. A)	LKKIIGQVREQAEHLKTAV	Tyrosine kinase binding motif
SIV/CPZ (US)	LKKIIGQIRDQAEHLKTAV	Tyrosine kinase binding motif
HIV-2 (cl. A)	LKNQISRIREQANTZETIV	No motif found
SIV/SMM	LKNQISRIREQANSZETIV	No motif found
SIV/AGM (VER)	LKEIIGKIRDDCQYTETAV	Tyrosine kinase binding motif
SIV/AGM (SAB)	LKEIIGQIRDDAERLETAV	Tyrosine kinase binding motif
SIV/SYK	LKEAISQIRDDVTHLQTAV	No motif found
SIV/L'HOEST	LKKIIGQVRDQAEHLKTAV	Tyrosine kinase binding motif
PFV	ARPASLRPRWHKPSTVLKVLNPR	Protein kinase C binding motif
SV40 (T-Ag)	MPKKKRKVEDPGT	Tyrosine kinase binding motif

The foamy virus is different from other retroviruses, its NLS is more likely phosphorylated by protein kinase C (pattern: [ST]-x-[RK]). Hypothetically, viral NLSs could be phosphorylated by different cytoplasmic kinases. That may play a crucial role in nuclear import and chromosomal integration of retroviruses. Kinases could trigger a cascade of intracellular events, which via protein phosphorylation, may transmit the signal (NLS) to the nucleus (Gunby *et al*., 2007). In case of HBV, it was shown that phosphorylation promotes exposure of a NLS in the COOH-terminal part of core protein which allows core binding to the NPCs by the importin- (kariopherin-) mediated pathway (Kann *et al*., 1999). Moreover, the cellular kinase binding motifs (PxxP and RR) were found in the HIV-1 Nef protein. Those motifs were conceded to be indispensable for producer-cell-dependent enhancement of viral entry (Tokunaga *et al*., 1999).

Regulation by phosphorylation was found in the retrotransposons. These are closely related to the retroviruses in terms of the genome integration mechanisms. Prior to integration, transposon-encoded proteins may be directly targeted to diminish transposition activity. Retrotransposase phosphorylation leads to negative regulation of P-element excision in *Drosophila* (Rio *et al*., 1991; Martin and Garfinkel, 2003).

Different kinases mediate processes of phosphorylation. They contribute to different cellular events: from mechanisms of apoptosis (Vakifahmetoglu and Zhivotovsky, 2009) to the RNA export and splicing (Okaga and Ye, 2009; Feng *et al*., 2008). As a result, viral NLSs could have a dual role. On one hand, they serve as signals for NPC proteins recognition and binding. On the other hand, they contain KBMs for interaction with cellular kinases. It is thus tempting to establish a link between type of kinase, NLSs and proteins of NPC (TR-SRs etc).

The regulation of HIV-1 IN with host cell proteins (TR-SR2) is a new route in antiviral drug generation. Potential chemical compounds with the affinity to the viral NLS would preferentially bind to it and block its function. As was already shown, the HIV matrix protein NLS was successfully targeted by arylene bis(methylketone) compounds to prevent nuclear import of HIV PIC (Dubrovsky *et al*., 1995). From this point of view one could regulate viral integrase activity and block it by disrupting the transportin-integrase interaction via changing TR-SR2 affinity to HIV-1 IN. From this point of view the IN NLS could become a novel target for antiviral therapeutic intervention. In this context, reconsidering the processes of nuclear import opens new possibilities for treatment of viral infection by disrupting those links and trigger further experimental studies on possible PIC and NPC interactions.

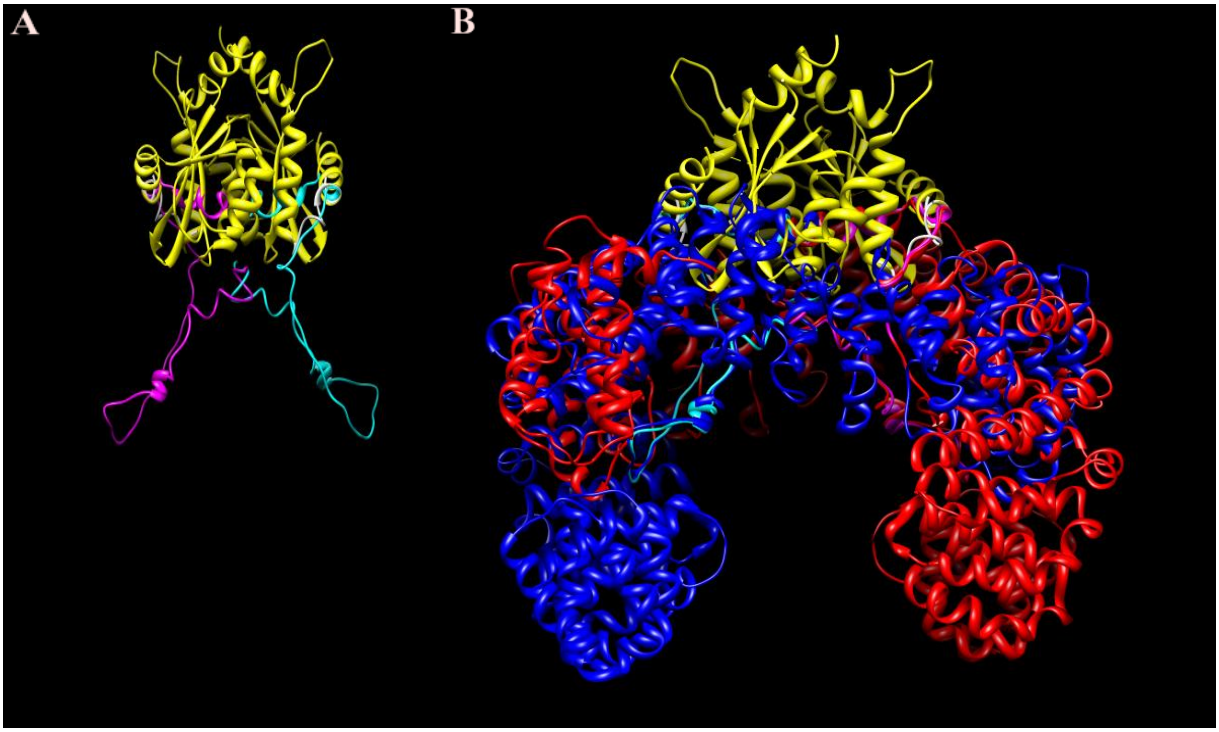


Figure 1.7: Scheme of the integrase - transportin complex. (A) integrase - H8-loop (HIV-1 integrase-yellow, NLSs-white, H8-loops-magenta), (B) integrase - TR-SR2s. The TR-SRs are blue and red respectively. The pictures were generated using the Chimera software.

1.5 Conclusions

Not all viruses enter the nucleus to integrate their genomes into the host-cell chromosome. However, for retroviruses this stage in the replication cycle makes them different from

other viruses. My study investigates and suggests a specific route of transportin-mediated nuclear import for these viruses. Interaction of HIV-1 with the TR-SR2 and probably with other cellular factors makes nuclear import very flexible (different transporters recruitment) and deals with a bulky process (complex PIC composition) at the same time **(Figure 1.7)**.

Probably uniqueness of retroviral nuclear import is also specified by particular cellular transporters such as TR-SR2. These transporters will interact exclusively with retroviral PICs. By natural selection, cellular NLSs were mimicked by retroviruses in order to get their genomes into the nucleus. It is important to understand this type of retroviral nuclear import as a part of retroviral speciation during the course of their evolution and for therapeutic applications.

Chapter 2

2.1 Overview

Retroviral vectors are potent tools for gene delivery and different research applications. To accomplish gene transfer task successfully, they must efficiently transduce different cell cultures at different phases of cell cycle. However, very promising retroviral vectors based on the foamy viral (FV) backbone are lacking ability to efficiently transduce quiescent cells. This phenomenon might be explained as inability of foamy viruses to form a pre-integration complex (PIC) for internuclear transport of viral DNA, which is the case for lentiviruses (HIV). During reverse transcription the central polypurine tract (cPPT) is served as a primer for plus strand synthesis and believed to be crucial for PIC formation. Here, I compare the transduction frequencies of PPT modified FV vectors with lentiviral vectors in quiescent and dividing adenocarcinomic human alveolar basal epithelial cells (A549) by several methods. I could not determine if mitosis facilitated nuclear entry of FV vectors, since cell-free vector preparations contained long terminal repeat circles, precluding their use as nuclear markers. In contrast to lentiviral vectors, FV vectors were not able to efficiently transduce nondividing cell in my hands. Despite the previous findings, which support the use of FV vectors as a safe and efficient alternative to lentiviral vectors, there is still a big limitation in terms of FV-based vector efficient gene transfer in quiescent cells.

2.2 A molecular biology challenge

Retroviral vectors are efficient gene transfer vehicles that deliver different transgenes by precise integration into a genome of host cells. This remarkable ability has been widely exploited for many gene applications, including gene therapy.

In order to replicate, all retroviruses pursue very complex process, which is known as a reverse transcription (RT). During this event, they will transcribe their single-stranded RNA genomes into double-stranded DNA intermediate, in order to successfully integrate into a host genome. The result of reverse transcription is linear DNA with a long terminal repeats

(LTRs) at each extremity. HIV-1 **(Figure 2.1 A)** has evolved a complex reverse transcription strategy including a strand displacement event controlled by the central polypurine tract (cPPT) for plus-strand initiation (priming) and the central termination sequence (CTS) (Charneau and Clavel, 1991; Charneau *et al.*, 1992; Charneau *et al.*, 1994; Lavigne *et al.*, 1997).

Reverse transcription produces a linear DNA molecule bearing in its center a stable triple structure, known as central DNA flap element, because of the RT complex termination at the CTS site. This flap element is believed to be essential for preintegration cormplex (PIC) formation, which is composed of both viral and cellular proteins (Bowerman *et al.*, 1989; Brown *et al.*, 1989; Shityakov *et al.*, 2010) and presumably possessed the ability to cross the nuclear membrane in order to enter the nucleus **(Figure 2.1 ; Figure 2.2 A)**. It was reported previously that matrix protein contains the nuclear localization signals (NLS) for PIC-mediated nuclear import (Bukrinsky *et al.*, 1993; van Schwedler *et al.*, 1994). However, the main mechanisms of these processes are still unknown. Recent reports have indicated that integrase (IN) was located in the nucleus (Ikeda *et al.*, 2004; Lu *et al.*, 2004). These data are in accordance with the hypothesis that IN regulates the nuclear import of the viral PIC. Some cellular factors, such as Transportin-SR2, are also might be involved in the PIC formation (Shityakov *et al.*, 2010).

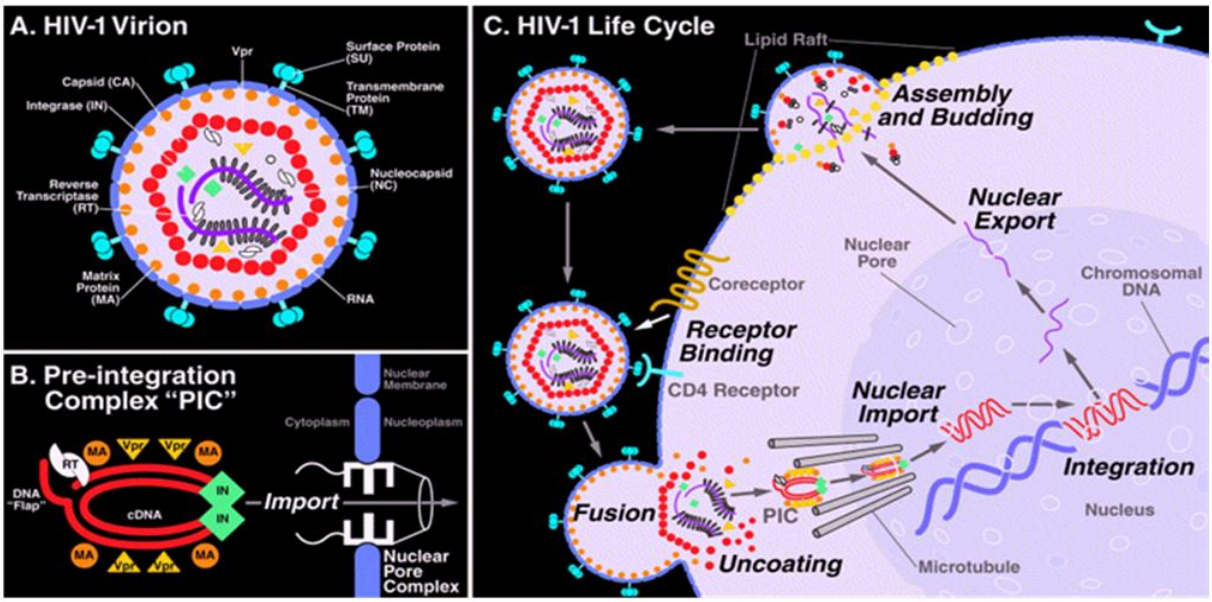


Figure 2.1: (A) mature and fully assembled retroviral virion. (B) The complex organization of lentiviral preintegration complex (PIC). It includes the viral integrase (IN), reverse transcriptase (RT), matrix part (MA) of Gag and some auxiliary proteins, such as Vpr. The PIC

exceeds more than two times the size of the nuclear pore central channel within the nuclear pore complex yet is able to successfully negotiate into the nucleus. ***(C)*** *Briefly, the HIV life cycle comprises cognate co-receptor initiating fusion, uncoating, reverse transcription, nuclear import, virion production and budding (figure modified from Sherman and Greene, 2002).*

However, there are some retroviruses such as foamy viruses (FV) which have a replication strategy and probably PIC formation mechanisms different from those in lentiviruses. FVs belong to *Spumavirinae* retroviral subfamily. FVs are non-pathogenic retroviral species and endemic to a number of mammalians such as nonhuman primates, cats and cows (Hooks and Gibbs, 1975). These viruses were isolated in humans and described as prototype foamy viruses (PFV). They turned out to be highly homologous to monkey virus initially (Brown *et al.,* 1978., Herchenroder *et al.,* 1994; Nemo *et al.,* 1978; Schweizer *et al.,* 1995). More recently, sequence comparisons of the integrase gene between the prototype foamy isolate (Achong *et al.,* 1971), and those of simian isolate obtained from the four chimpanzee subspecies (SFVcpz), established the link that that PFV is most closely related to the FV from *Pantroglodytes schweinfurthi*, the chimpanzee species present in Kenya (Switzer *et al.,* 2004).

The foamy viruses have a very broad host range and infect a wide variety of different permissive cell lines, including those as the fibroblastoid, epitheloid, and lymphatic lineages (Mikovits *et al.,* 1996). At the present time, no resistant cells have been detected for foamy viral infection. As no any harmful effect for cells caused by FV infection was detected *in vivo*, it leads to a massive cell death via apoptosis *in vitro* (Meiering *et al.,* 2003). The mechanism of this phenomenon is still remained to be determined.

FVs seem to have a different outcome of plus-strand priming event than lentiviruses. Foamy viral reverse transcription model was proposed by Li and co-authors in 1993. In the FVs, the strong stop plus-strand DNA is displaced by DNA elongating from cPPT. This DNA can further be transferred to the 3' extremity of the minus-strand DNA or serve as a template for synthesis of a double-stranded LTR. The RNA sequence of the cPPT is resistant to the retroviral RNAse H activity. Notwithstanding, the cPPT will degrade after completion of the minus-strand cDNA synthesis resulting in single-stranded gap structure formation (Huber and Richardson, 1990; Luo *et al.,* 1990; Pullen el al., 1993). The gap was then shown to be 120 nucleotides long and located on the plus DNA strand (Kupiec *et al.,* 1988; 1991) **(Figure 2.2 B).**

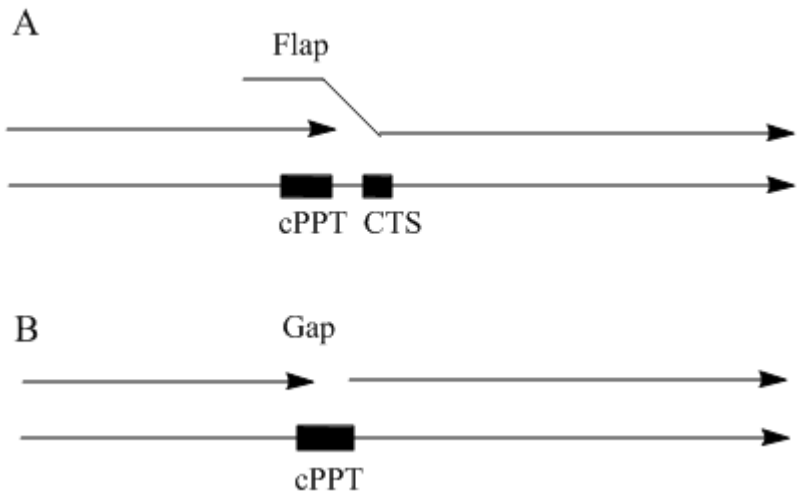


Figure 2.2: Different reverse transcription outcomes in lenti- (A) and foamy viruses (B). cPPT-CTS abbreviation stands for the central polypurine tract and the central termination sequence, respectively (see text for details).

There are several FV-based vectors for a gene therapy that have been developed recently (Bieniasz *et al.*, 1997; Heikelein *et al.*, 2002). These constructs have several desirable properties in comparison to lentiviral vectors. The main characteristics of foamy viral vectors are: 1) safety, FVs have not been linked to any established pathology (Linial, 2000; Meiering and Linial, 2001), hence FV-associated disease of primates will not evolve in infected humans (Heneine *et al.*, 1998; Meiering and Linial, 2001); 2) FV vectors have a wide tissue tropism (Hill *et al.*, 1999; Hooks *et al.*, 1975; Mergia *et al.*, 1996; Russell and Miller, 1996); 3) the viral particles are stable and resilient to centrifugation at high velocities (Vassilopoulos *et al.*, 1998); 4) foamy viral packaging efficacy for foreign DNA is up to 9 kb and will provide valuable tool to transfer big genetic elements and therapeutic transgenes (Trobridge *et al.*, 2002); 5) replication-defective FV vectors can be produced at titres sufficient for *ex vivo* gene therapy applications (Trobridge *et al.*, 2002; Trobridge *et al.*, 1998).

However, very promising foamy viral vectors are more likely not able to efficiently transduce quiescent cells. This phenomenon might be explained by means of the FV deficiency to form a pre-integration complex (PIC) for internuclear transport of viral DNA, which is present in lentiviruses (HIV).

Finally, the role of cPPT in nuclear import, and the mechanism by which it functions are still remained to be clarified. To pursue this goal, I compared the transduction frequencies of cPPT mutated FV vectors with lentiviral ones in quiescent and dividing adenocarcinomic human alveolar basal epithelial cells (A549). My data indicates that FV vectors are

not able to transduce nondividing cell efficiently. My comparison of these two classes of retroviral vectors suggests that FV vectors may be not the best choice for *ex vivo* stem cell gene therapy. Hence, they may not work well in the cell types that remain quiescent, such as postmitotic myotubes or neurons. Further studies will be needed to directly compare these two classes of vectors in preclinical gene therapy experiments and to define the nature of the stable FV vector transduction intermediate in quiescent cells.

2.3 Materials and methods

2.3.1 Materials and solutions

Restriction enzymes were from the Fermentas and the NEB (New England Bio-labs, GB). "QIAprep Spin Miniprep Kit" and "Qiagen Plasmid Maxi Kit" were from Qiagen (CA, USA). Tryptone peptone, yeast extract and agar were from Difco (Becton Dickinson, USA). Luria broth (LB): 1% (w/v), tryptone peptone, 0.5 % (w/v) yeast extract, 1% NaCl; auto-claved 15 minutes at 121 °C. TE buffer: 10 mM Tris-HCl, pH 8.0, 1mM EDTA, 50 x TAE buffer (1 liter). Aphidicolin from Sigma-Aldrich is a potent antiviral and antimitotic agent and DNA polymerase inhibitor, used to stop cell cycle at G_1/S phase. The drug was dissolve in pure DMSO to reach 10 mg/ml of its concentration. Initially, Aphidicolin was extracted from *Nigrospora sphaerica*.

2.3.2 General molecular genetics methods

Agarose gel preparation

Agarose gels were prepared by dissolving an agarose (Bio-Rad Laboratories, UK) in 1x TAE buffer and heating with a microwave oven. 0.5 µg/ml of ethidium bromide were added prior to gel casting. Than the gel was poured and run on an agarose gel apparatus. 0.3-0.5 µg of DNA standards were loaded and often used as a reference to give an approximate estimation of DNA samples.

Preparation of competent cells

A single colony of the Top-10 *E. coli* bacteria cells (Stratagene, CA, USA) was picked from LB-agar plate and used to inoculate 10 ml of LB. Bacteria were grown over-night, the culture was diluted in 190 ml of fresh LB and incubated at 37 °C until the optical

density reached 0.5 (at 600 nm). Bacteria were harvested by centrifugation at 4,000 x g for 10 minutes at 4 °C. The bacterial pellet was resuspended in 64 ml of Transformation solution (recipe). The bacteria were stored at -80°C.

PCR amplification of DNA inserts

To amply specific regions of DNA inserts, PCR was performed incubating 10 ng DNA-plasmid as template in 50 µl of 20 mM Tris-HCl pH 8.8, 10 mM KCl, 2.0 mM $MgSO_4$, 10 mM $(NH_4)SO_4$, 0.1% Triton X-100, 0.1 mg/ml nuclease free BSA, 1.0 µM each oligonucleotide, 200 µM dNTP, 2.5 U Pwo DNA polymerase. The PCR reaction mixtures were subjected to 35 temperature cycles in a programmable thermal cycler.

Melting-, annealing- and elongation temperatures were adjusted according to the features of the templates and the primers. To facilitate the subsequent subcloning of the PCR products, the forward and the reverse primers were provided with restriction sites at their 5' ends (see specific section). The appropriate plasmid was used as a template to amplify the HIV cPPT such as the pWPXL vector.

The pWPXL is the lentiviral Cre/Lox plasmid for mammalian constitutive transgene expression. pWPXL contains the EF-1α promoter and intron that promotes high expression of the gene of interest, hence the RNA will easily be spliced and go more efficiently out of the nucleus. The loxP site in the 3'LTR (long terminal repeat) is duplicated to the 5'LTR during reverse transcription in the target cells. This allows for removal of an integrated provirus by cre-element (cardiovirus cis-acting replication).

Restriction and ligation

DNA (vectors and inserts) were cut with 5 U/µg of the appropriate restriction enzyme in the buffer supplied with each enzyme (Fermantas or NEB). After restriction, enzymes were usually inactivated by incubation at 65-75 °C for 10-20 minutes, according to the manufacturer's instructions and loaded on 1% agarose gel. The bands of interest were cut out of the gel with a sterile scalpel and the DNA was extracted from the gel with extraction kit (Qiagen, CA, USA), according to the manufacturer's instructions. DNA was eluted in the Tris-HCl buffer pH 8.0. To ligate a vector with an insert, about 100 ng of the particular vector construct and about 3-fold molar amounts of insert were incubated with 1U of T_4 DNA ligase in the T_4 DNA ligase buffer for 10 minutes at room temperature.

Transformation of the competent bacteria cells

The DNA plasmid of interest (10 ng DNA of ligation reaction) was added to 100-150 µl of the Top-10 competent cell (see previous sections) previously thawed on ice. After gentle mixing, the cells were left on ice for 30 minutes and that subjected to the heat shock in 42 °C water for 90 seconds. Bacteria were placed on the LB agar plates containing the appropriate selective antibiotic marker and incubated overnight at 37 °C. The next day, isolated bacterial colonies were picked with sharp needle tip and used to inoculate the 2-3 ml of the 1 x LB containing appropriate antibiotic. They were incubated at 37 °C overnight. For storage purposes, 700 µl of bacterial cultures were added to 300 µl of 50% sterile glycerol and store at -80 °C indefinitely.

Construction of expression vectors

Standard techniques of molecular cloning were used for constructs generation: pMD9-cPPT (HIV)-CTS (HIV), pMD9-cPPT (PFV)-CTS (HIV), pMD9-cPPT (HIV), pMD9-CTS (HIV) **(Figure 2.3)**. Vector viruses were abbreviated with the plasmid name lacking the 'p' and 'pco' for codon-optimized plasmid. To create all my pMD9-based vectors, I have to construct the intermediate pUC19-based subclone (pUC19-cPPT (PFV)) in order to use appropriate restriction sites. The restriction enzymes for all the constructs were MfeI/StuI and EcoRI/KpnI accordingly.

To generate the pUC19-cPPT (HIV)-CTS (HIV) subclone, I initially amplified the cPPT (HIV)-CTS (HIV) sequence using the pWPXL (HIV virus backbone) vector as a template (forward primer: 5'–TATACAATTGATGGCAGTATCCAC–3'; reverse primer: 5'–TATAAGGCCTCTGTCCCTGTAATAAACC–3') and than ligated MfeI/StuI digested cPPT (HIV)-CTS (HIV) fragment with MfeI/StuI digested pUC19-cPPT (PFV) fragment.

To generate the final pMD9-cPPT (HIV)-CTS (HIV) plasmid, I ligated the cPPT (HIV)-CTS (HIV) fragment from pUC19 subclone with the pMD9 vector using EcoRI/KpnI restriction enzymes.

To generate the pUC19-cPPT (PFV)-CTS (HIV) subclone, I initially amplified the CTS (HIV) sequence using the pWPXL vector as a template (forward primer: 5'–TATACAATTGCAGGAGAGGGATTGGGGGGTACAGTGCAG–3'; reverse primer: 5'–TATAAGGCCTCTGTCCCTGTAATAAACC – 3') and than ligated MfeI/StuI digested CTS (HIV) fragment with MfeI/StuI digested pUC19-cPPT (PFV) fragment.

To generate the final pMD9-cPPT (HIV)-CTS (HIV) plasmid, I ligated the cPPT (PFV)-CTS (HIV) fragment from pUC19 subclone with the pMD9 vector using EcoRI/KpnI restriction enzymes.

To generate the pUC19-cPPT (HIV) subclone, I initially amplified the HIV central polypurine tract (cPPT) sequence using the pWPXL vector as a template in polymerase chain reaction (forward primer: 5'–TATACAATTGATGGCAGTATCCAC–3'; reverse primer: 5'–TATAAGGCCTGTAATTTGTTTTTGTAATTCT – 3') and than ligated MfeI/StuI digested cPPT (HIV) fragment with MfeI/StuI digested pUC19-cPPT (PFV) fragment.

To generate the final pMD9-cPPT (HIV) plasmid, I ligated the cPPT (HIV) fragment from pUC19 subclone with the pMD9 vector using EcoRI/KpnI restriction enzymes.

To generate the pUC19-CTS (HIV) subclone, I initially amplified the HIV central termination sequence (CTS) using the pWPXL vector as a template in polymerase chain reaction (forward primer: 5'–TATATACAATTGGGGGGTACAGTGCAGGGG–3'; reverse primer: 5'–TATATAAGGCCTTCCCTGTAAACCCGAAAATTTTG–3') and than ligated MfeI/StuI digested CTS (HIV) fragment with MfeI/StuI digested pUC19-cPPT (PFV) fragment.

To generate the final pMD9-CTS (HIV) plasmid, I ligated the CTS (HIV) insert from pUC19 subclone with the pMD9 vector using EcoRI/KpnI restriction enzymes.

Finally, the distance between the cPPT (PFV) and CTS (HIV) in the pMD9-cPPT (PFV)-CTS (HIV) plasmid is 87 base pairs. The distance between the cPPT (HIV) and CTS (HIV) in the pMD9-cPPT (HIV)-CTS (HIV) is 77 base pairs.

The constructs were checked by nucleotide sequencing for possible mutations of the inserts. All generated vectors contained the gene encoding enhanced green fluorescent protein (GFP) reporter under the control of a constitutively active heterologous U3 promoter to enable the quantification of vector transfer rates.

Table 2.1 Main characteristics of the inserts using for the construction of modified expression vectors (see text for details).

Name of inserts	Length (bp)	Sequence of inserts
cPPT (PFV)	9	AGGAGAGGG
cPPT (HIV)	33	ATCCACAATTTTAAAAGAAAAGGGGGGATTGGG
CTS (HIV)	16	AAAAATTCAAAATTTT

2.3.3 General cell biology methods

Cell cultures

Human embryonic kidney cells (HEK 293T) and adenocarcinomic human alveolar basal epithelial cells (A549) were purchased from American tissue type collection (VA, USA). Modified Eagles medium (MEM), Dulbecco's modified Eagles medium (DMEM), non essential amino acids (NEA), penicillin, streptomycin, trypsin/EDTA and L-glutamine were from Gibco (NY, USA).

HT1080 cells were grown in MEM supplemented with 10% FCS, 4mM L-glutamyne, 0.1 mM NEA, 100 U/ml penicillin and 100 µg/ml streptomycin. All cells were grown under a controlled atmosphere in the presence of 5 % CO_2 at 37 °C. Cells were grown in a flask until 80-100% confluence. Subsequently, the medium was removed and trypsin/EDTA solution (0.05% trypsin, 0.02% EDTA) was added for 2-5 minutes. Finally, the medium was added back to block a protease action and the cells were collected in a plastic flask in proportion: 0.5 ml of cells per 9.5 ml of medium.

Cell cultures transfection, transduction and supernatant purification

The retroviral system for gene therapy comprises two principal elements to be fully functional: truncated viral backbone and viral packaging plasmids. The PFV system includes the elements such as: 1) pMD9 (Cas I, II, cPPT) 2) pcoGAG 3) pcoPOL 4) pcoENV. The negative control system includes the elements such as: 1) pMD9 (without pcoENV) 2) pcDNA (mock). The HIV system includes the structural elements such as: 1) pWPXL (ψ, Pol, cPPT-CTS) 2) psPAX2 (Gag) 3) pMD2.G (VSV-G, pseudotyping).

On the first day, HEK 293T cells (2 x 10^6) were seeded into 6.0 cm dishes and than transfected with the DNA plasmid using polyethylenimine transfection (PEI) reagent (Polyscience). The transfection mixtures contained appropriate volumes: 1) 2.5 µg pWPXL + 1.9 µg psPAX2 + 0.8 µg pMD2.G (VSV-G); 2) 2.5µg pMD9 +1.25µg pcoGAG + 0.5µg pcoPOL + 0.5µg pcoENV; 3) 2.5 µg pMD9 + 1.25 µg pcoGAG + 0.5 µg pcoPOL + 0.5 µg pcDNA; 4) 0.5 µg pcDNA; 5) 2.5 µg pMD9-cPPT (HIV)-CTS (HIV) + 1.25 µg pcoGAG + 0.5 µg pcoPOL + 0.5 µg pcoENV; 6) 2.5 µg pMD9-cPPT (PFV)-CTS (HIV) + 1.25 µg pcoGAG + 0.5 µg pcoPOL + 0.5 µg pcoENV; 7) 2.5 µg pMD9-cPPT (HIV) + 1.25 µg pcoGAG + 0.5 µg pcoPOL + 0.5 µg pcoENV; 8) 2.5 µg pMD9-CTS (HIV) + 1.25 µg pcoGAG + 0.5 µg pcoPOL + 0.5 µg pcoENV. Subsequently, 245.25µl of pure MEM medium and 250 µl of PEI mixture (18 µl PEI in 232 µl of pure MEM medium) were added to all transfection mixtures. The transfection mixture (500 µl) was added in drops to the plates and incubated overnight. All the mixtures were incubated for 15-30' at room temperature.

On the second day the sodium butyrate was added in the morning: 60 µl per plate and after 6 hrs medium was exchanged. Than 3.0 ml of pre-warm 10% FSC MEM medium was added to the mixture. A549 cells were prepared in advance in 12-well plate (V = 500 µl, c = 200 000 cells/well).

On the third day the supernatant from HEK 293T cells was produced using 0.45 µm filter and 2.0 ml syringe. A549 cells were infected and incubated for 3 days. The vector transfer assays were done at least three times with all plasmid preparations.

After three days, the cells were prepared for FACS (fluorescence-activated cell sorting) analysis using 2.0 ml FACS tubes filled with the FACS buffer in each tube (0.1 % of BSA in PBS + sodium acetate). Cells were detached from 12-well plate and transferred to 2.0 ml FACS test-tube. Prior to FACS analysis, the cells were centrifuged at 1200 rotations per minute at 20 °C for 5 minutes. Finally, the expression of enhanced green fluorescent protein (eGFP) was monitored and measured.

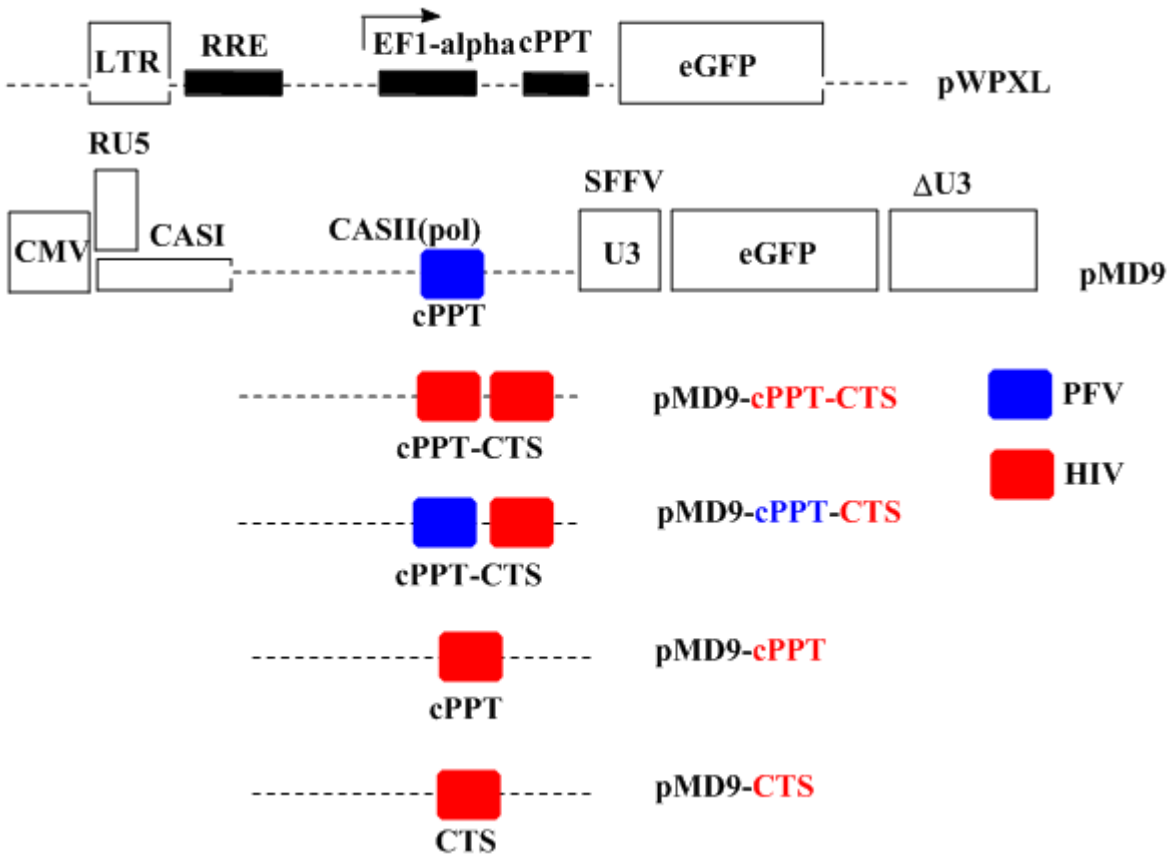


Figure 2.3: The structural elements of pWPXL and pMD9 parental plasmid backbones are shown. Foamy viral central polypurine tract (cPPT) was replaced with the HIV cPPT in different variations. Central termination sequence (CTS) was deleted or inserted in accordance

with the experimental scheme shown above. The details, concerning the construction of corresponding vectors, are described in Materials and Methods section.

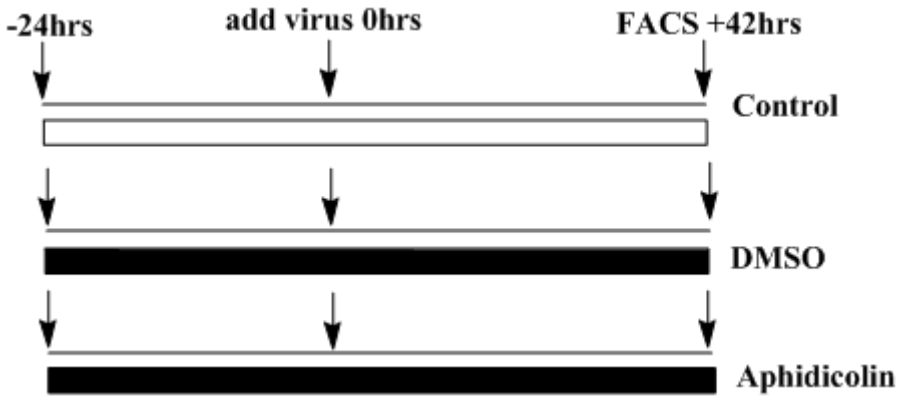


Figure 2.4: The timescale representation of the transduction experiments with Aphidicolin substance to stop a cell cycle (G_1/S phase). DMSO was added to observe its effect upon the A549 cell-line in pure solution, and as a solvent for the drug.

2.4 Results and discussion

Here, I generated the cPPT mutants on the basis of the pMD9 retroviral system. The following mutants are 1) pMD9-cPPT (HIV)-CTS (HIV); 2) pMD9-cPPT (PFV)-CTS (HIV); 3) pMD9-cPPT (HIV); 4) pMD9-CTS (HIV). Further, I compared the cell cycle requirements for efficient cell transduction by pMD9- and pWPXL-based vectors expressing the same transgene (eGFP) under control of different promoters (LTR U3/EF1-α) in A549 mammalian cells. To study the cPPT-CTS (HIV) effects upon the foamy viral replication, pMD9 modified vectors were used in order to increase vector infectivity in growth – arrested (G_1/S) permissive cells. However, the previous studies have already focused on the cell cycle effects upon retroviral infection; my data are the first direct attempt to analyze the role of cPPT modifications from two classes of retroviruses: lentiviruses, and spumaviruses.

It appears that after disassembly in the cytoplasm, some retroviral NLS sequences are recognized by host-cell import factors such as transportins that mediate nuclear targeting of PICs via the nuclear pore complex (Shityakov *et al.*, 2010). Several such sequences are present in HIV-1, most significantly in the virus integrase (IN) (Bukrinsky *et al.*, 1993; Bouyac-Bertoia *et al.*, 2001). Foamy viral Gag and Pol proteins, including IN, contain NLS sequences

(Imrich *et al.*, 2000) that, in combination, might play some role in the transduction ability, however, not solely enough to efficiently infect arrested cells. The lower efficiency in comparison with lentiviral vectors could be due to the PFV NLS being partly occluded in the PIC (Saïb *et al.*, 1997) showed that PFV generates 2-LTR circles in nondividing cells. It is possible that there are further blocks to PFV integration after nuclear translocation of the DNA.

FV vectors have distinct requirements for transduction in that a partial cell cycle including mitosis is ultimately required for transgene expression, but a transduction intermediate is stable and persists in nondividing cells. My finding, that mitosis is a critical phase in the cell cycle for FV transduction, confirms those of Bieniasz *et al.*, who showed that G1/S or G2 arrest blocked FV protein expression in infected cells.

Retroviruses have a tendency to package some cellular proteins non-specifically, and DNA ligase activity has been demonstrated in virions (Temin *et al.*, 1972), so these steps may occur in extracellular FV virions that contain cDNAs or before virions are released from cells. Some studies detected FV LTR circles by PCR in Aphidicolin-arrested but not serum-deprived cultures (Saïb *et al.*, 1997). This issue needs to be addressed further with additional experiments using other methods. Probably there is a possibility that FV vectors use distinct mechanisms for nuclear import, as both the viral genome and Gag proteins accumulate near the centrosome (Saïb *et al.*, 1997).

My results conflict with a prior study that concluded that simian FV vectors can efficiently transduce Aphidicolin-treated cells (Mergia *et al.*, 2001). Thus, while I cannot rule out strain-specific differences of the vector systems used, it is very probable that simian FV vectors require mitosis for efficient transduction. While FV vectors require breakdown of nuclear membrane for nuclear entry and thus transduction (Roe *et al.*, 1993), lentiviral vectors can enter nucleus using nuclear localization signals in nondividing cells (Bukrinsky *et al.*, 1993; Popov *et al.*, 1998).

Three conditions were analyzed in the experiment to test the generated mutants to efficiently transduce non-dividing cells (G_1/S phase of cell cycle): 1) in pure medium; 2) in presence of DMSO; 3) in presence of Aphidicolin dissolved in sterile DMSO. DMSO-treated cells have shown the decrease of transduction rates for pMD9-based vectors with and without cPPT modifications in comparison to the control **(Figure 2.5 A, B).** These results indicate that productive infection was observed only if target cells were allowed to pass through mitosis.

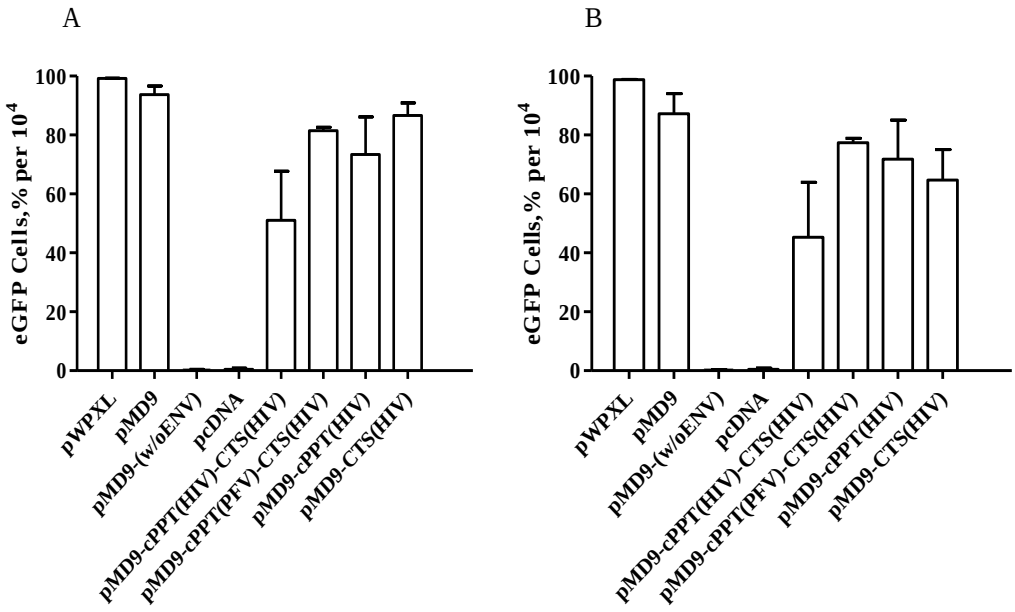
One point to note about the data is that different promoters directed the eGFP trangene expression in the vectors. In this regards it is hard to emphasize the levels of trans-

gene expression under control of the different promoters. In my experiments I used the pMD9 vector with the eGFP transgene driven by the LTR U3 and pWPXL vector with the eGFP transgene driven by the EF1-α promoter.

There was a huge difference in the relative transduction efficiencies of the vectors, where the cells were arrested in 24 hrs before they were transduced **(Figure 2.4)**, and the transduction efficiency of pWPXL was almost 4000 fold higher than that of the pMD9 vectors **(Figure 2.5 C)**.

Finally, I was not able to successfully rescue the cPPT mutated pMD9-based vectors from the specific conditions in nondividing cells, which restrain the ability of foamy viral vectors to efficiently infect quiescent cells. This fact indicates the incompatibility of HIV cPPT and FV cPPT, hence these structures cannot be interchanged without the severe impairment of retroviral reverse transcription and loss of overall infectivity.

The cell cycle dependence of foamy virus infection shows that vectors based on foamy virus genomes and packaging proteins may not be useful for targeting nondividing cells. Nonetheless, the neurotropism of foamy viruses is worth testing in animal models, but further studies will be needed to directly compare these two classes of vectors in preclinical gene therapy experiments and to define the nature of the stable FV vector transduction intermediate in quiescent cells.



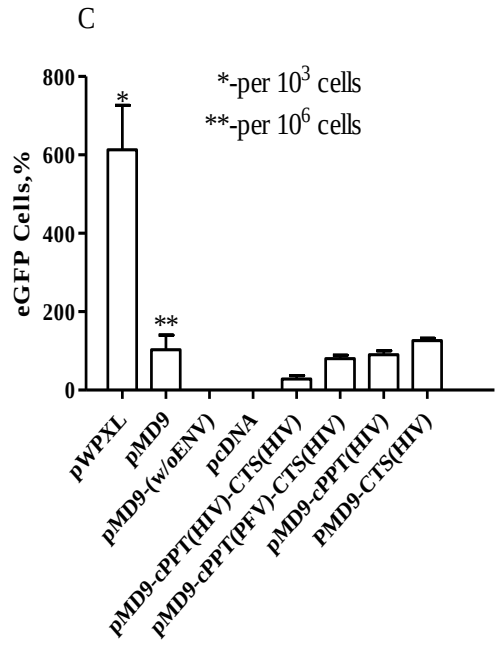


Figure 2.5: The summarized statistics of retroviral replication efficiency in A549 cells. pMD9 mutants were tested to efficiently transduce non-dividing cells (G_1/S phase of cell cycle). Three conditions were analyzed in the experiment: pure medium (A); DMSO (B); Aphidicolin + DMSO (C).

2.5 Conclusions

In my experiments, modified foamy viral vectors showed a reduction in transduction rates of dividing A549 cells. PFV vectors are able to infect G1/S growth-arrested A 549 cells but at extremely low titers in comparison to lentiviral vectors. Finally, the HIV-1 cPPT cannot be substituted for the PFV cPPT without loss of functionality, probably because of the cognate RT enzyme is required for recognition of PPTs and CTSs sequences.

From my data, it is clear that foamy viruses are not able to efficiently infect non-dividing cells probably because of low their integration efficiency that of lentiviruses unless cells are undergoing mitosis. The results presented here are in accordance with the findings of Trobridge and Russell (2004), but are contrary to those published by Mergia *et al.*, 2001. Hence, this data spark the controversy that surrounds the question of whether foamy viral vectors derived are really to be effective in gene therapy targeted to nondividing cells.

Chapter 3

3.1 Overview

Indinavir (*Crivaxan*®) is a potent inhibitor of the HIV (human immunodeficiency virus) protease. This enzyme has an important role in viral replication and is considered to be very attractive target for new antiretroviral drugs. However, it becomes less effective due to highly resistant new viral strains of HIV, which have multiple mutations in their proteases. For this reason, I used a lead expansion method to create a new set of compounds with a new mode of action to protease binding site. 1300 compounds chemically diverse from the initial hit were generated and screened to determine their ability to interact with protease and establish their QSAR properties. Further computational analyses revealed one unique compound with different protease binding ability from the initial hit and its role for possible new class of protease inhibitors is discussed in this chapter.

3.2 The strategy

The HIV-1 (human immunodeficiency virus type 1) protease is a C2 symmetric and an aspartic acid homodimeric enzyme, where Aspartate-25 plays a pivotal role in binding the substrate (York *et al.*, 1993). It differs from the well-known monomeric enzymes of this class, such as pepsin and renin, in that their structure is achieved by the assembly of two 99 amino acid polypeptides into a functional homodimer (Meek *et al.*, 1989; Wlodawer *et al.*, 1989).

The HIV-1 protease does not have a homologue in mammalian cells and has a function to cleave the Gag-Pol polyprotein precursor into shorter pieces to create the active protein components for viral packaging and maturation. This proteolysis occurs late in the viral life cycle and is essential for viral infectivity (Kohl *et al.*, 1988).

In vitro studies have shown that other substrates may serve as substrates of HIV protease, including calmodulin, fibronectin, cytoskeletal and sarcomeric proteins (Winslow and Otto, 1995).

Six regions (flaps, ear, cheek, eye, nose, whiskers) of the HIV protease are highly conserved across isolates of HIV-1 **(Figure 3.1)**. Some of these regions are associated with important characteristics of the enzyme: the substrate-binding region and the active site (Winslow and Otto, 1995). At the amino acid level, HIV-2 and SIV (simian immunodeficiency virus) vary ≈ 50 % from HIV-1. However, most of the amino acid variations occur outside the active site (Tomasselli *et al.*, 1990).

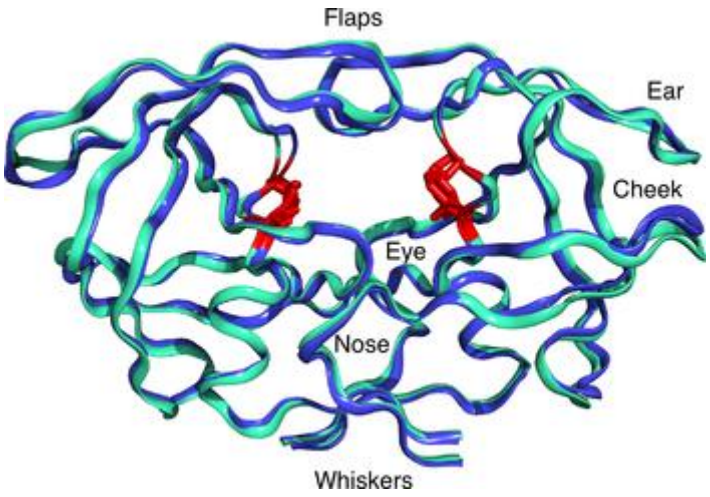


Figure 3.1: Crystallographic structures of the HIV-1 protease. The HIV-1 protease labelled according to its resemblance to an English bulldog (figure modified after Perryman et al., 2004). The blue and cyan-green ribbons depict the peptide backbone of a wild-type and a mutant structure, respectively.

The protease active site is located in the buried area (tunnel), where the two subunits meet each other. The active site lies between the identical subunits and has the characteristic Asp-Thr-Gly (Asp25, Thr26 and Gly27) sequence common to aspartic proteases. The two Asp25 residues (one from each chain) act as the catalytic residues. According to the mechanism for HIV PR protein cleavage proposed by Jaskolski *et al.*, water acts as a nucleophile, which acts in simultaneous conjunction with a well-placed aspartic acid to hydrolyze the scissile peptide bond. Inhibition of HIV-1 protease activity leads to production of non-infectious virions, which have the morphological features of immature particles (Kohl *et al.*, 1988).

As a step toward the development of drugs to combat HIV infection, numerous strategies have led to the design of a wide variety of HIV-1 protease inhibitors. Highly active antiretroviral therapy (HAART), using protease inhibitors (PI), is commonly used in management of HIV infection. These inhibitors are able to irreversibly bind the HIV-1 protease to block its function. Among these compounds, Indinavir (Crixivan, MK-639, or IDV) (Dorsey

et al., 1994; Vacca *et al.*, 1994) is a potent and selective protease inhibitor that had approval for AIDS therapy. But genotypic analyses of the viral populations during the course of protease inhibitor therapy had shown various mutations that can occur in as many as 20 amino acids within the protease gene (Condra *et al.*, 1996; Condra *et al.*, 1995) 6). 13.7 % of IDV failure were identified for 1021 new viral genotypes from HIV patients (Poveda *et al.*, 2007). In case of HIV-1 subtype C (African strain), it had been shown that the most common primary mutation observed in PI treated patients was L90M (Cane *et al.*, 2001), a main course in multi-PI resistance establishment (Kempf *et al.*, 2001). Even naturally occurring polymorphism, such as L89M and I93L, located in the hydrophobic core of the enzyme would possibly change the shape of the substrate-binding cleft and diminish the potency of IDV (Coman *et al.*, 2008). Only flexibility of the newly synthesized compounds seems to overcome this effect. Here, I describe the designing scheme for possible HIV-1 protease inhibitors using a lead expansion protocol with a pharmacophoric-shape similarity scoring function.

3.3 Computational methods

3.3.1 HIV-1 subtype C protease and Indinavir structures

The three dimensional structure of the HIV-1 protease - IDV complex (PDB code: 2R5P) was retrieved from Protein Data Bank at 2.3 Å RMSD resolution (Coman *et al.*, 2008). The VADAR (volume, area, dihedral angle reporter) server (Willard *et al.*, 2003), which is an improved version of the PROCHECK software, was used for stereochemical validation of HIV-1 protease. Altogether, 95 % of all residues were in φ-ψ core areas. Indinavir was obtained from the PubChem database (CID: 5362440). IDV initial refinement was performed by means of the MarvinSketch program provided by ChemAxon (Pirok *et al.*, 2006).

3.3.2 Protease active site detection

Protein binding surfaces could have very complicated and irregular structures. Atoms could form pockets, cavities and tunnels. Solvent molecules can get into these tunnels from outer environment and move through them. Buried shape and volume of such tunnels vary in time due to protein dynamics and kinetics. Here, I analyzed buried tunnels of the protease using the CAVER module, to identify certain atomic positions of hidden binding moieties (Petrek *et al.*, 2006).

Before the search procedure, 136 water molecules, chloride and sodium ions were removed from IDV-bound protease (liganded holo-structure) protein database file. All calculations were performed on 32612 grid points. Catalytic tunnel (key interaction site) was detected in the protease structure **(Figure 3.4 A)**. The tunnel coordinates to specify the starting point in x, y and z axis are 2.5, 6.5, -7.5 respectively. These coordinates were taken from the AutoDock on-line manual (http://autodock.scripps.edu/faqs-help/manual).

3.3.3 Compound library generation

Pharmacophoric - shape similarity scoring function was used for compound library generation. This is build-in function integrated in the Muse™ molecular design workflow to accelerate the identification and optimization of lead candidates. This function seeks molecules that have fairly different structures but similar 3D pharmacophores and shapes. It is appropriate for lead expansion – exploration around an initial hit. Indinavir compound was used as a reference molecule and all similarities were measured with respect to it. The reference structure was also used as the initial population for invention. In my case a default set of compounds was used if no any other structures were given. Such generation of multiple compounds with similar shape but different 3D structure organization would provide more interesting results in the invention and library screening afterwards. For lead expansion method, I did not specify any seed structures and preserved cores. 99 undesirable and 35 questionable substructures were excluded from to be present in invented structures. Structural ranking was done by means of the Pareto-Borda method to maximize pharmacophoric-shape similarity in preferential shape similarity specified range (from 0.35 to 0.50 score units).

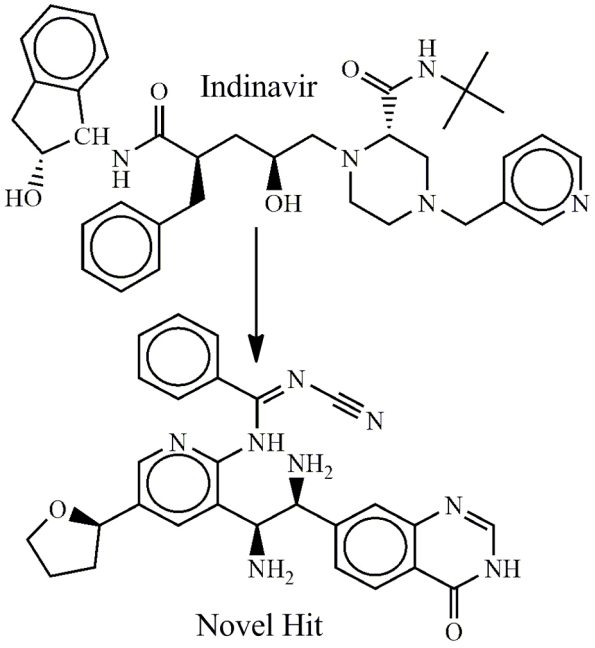


Figure 3.2: Chemical structure of IDV and its 'derivate'- novel hit (N'-cyano-N-{3-[(1S, 2S)-1,2-diamino-2-(4-oxo-3, 4-dihydroquinazolin-7-yl)ethyl]-5-[(2R)-oxolan-2-yl]pyridin-2-yl} benzenecarboximidamide).

3.3.4 ADME/Tox studies

Available decision tree algorithms such as: Benigni-Rossa rulebase, Verhaar scheme, Cramer rules and START (Structural Alerts for Reactivity in Toxtree) biodegradability were used for the analysis. The solutions are expressed in terms of the determinants of the (square) coefficient matrix and of matrices obtained from it by replacing one column by the vector of right hand sides of the equations.

Benigni-Rossa rulebase predicts the possibility of carcinogenicity and mutagenecity by discriminant analysis and structural rules (Cramer *et al.*, 1978). In the Toxtree implementation of this rulebase, the processing of a query chemical gives rise to limited number of different outcomes, namely: a) no structural alerts for carcinogenicity are recognised; b) one or more structural alerts (SAs) are recognised for genotoxic or non-genotoxic carcinogenicity; c) SAs relative to aromatic amines or unsaturated aldehydes are recognised, and the chemical goes through Quantitative Structure-Activity Relationship (QSAR) analysis, which may result in a negative or positive outcome. START biodegradability estimates biodegradation potential of the chemical compound based on structural alerts compiled from the Canadian Environmental Protection Agency.

3.3.5 Protein-ligand docking

The Molegro Virtual Docker (MVD) (Thomsen and Christensen, 2006) and the AutoDock software were implemented to analyze ligand interactions with the HIV-1 protease binding site. The Molegro Virtual Docker and the AutoDock (see appropriate section) combine a rapid energy evaluation through pre-calculated grids of affinity potentials with a variety of search algorithms to find appropriate binding positions.

The docking search algorithm (MolDock Optimizer) used in MVD is based on an evolutionary algorithm (Michalewicz, 1992; 2000). Evolutionary algorithms (EAs) are iterative optimization techniques inspired by Darwinian evolution theory. In EAs, the evolutionary process is simplified and thus it has very little in common with real world evolution. Nevertheless, during the last fifty years EAs have proved their worth as powerful optimization techniques that can assist or replace traditional techniques when these fail or are inadequate for the task to be solved.

Basically, an EA consists of a population of individuals (candidate solutions), which is exposed to random variation by means of variation operators, such as mutation and recombination. The individual being altered is often referred to as the parent and the resulting solution after modification is called the offspring. Sometimes more than one parent is used to create the offspring by recombination of solutions, which is also referred to as crossover. **Figure 3.3** below shows an outline of the evolutionary process taking place in EAs.

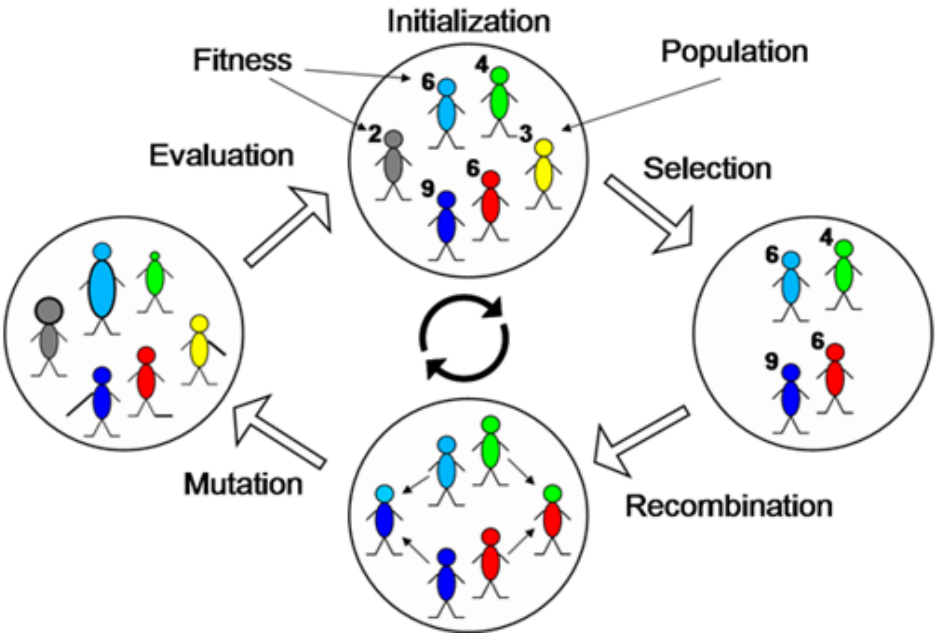


Figure 3.3: Outline of evolutionary algorithm.

In the Molegro Virtual Docker docking experiment I used MolDock scoring function, which is based on a piecewise linear potential and a re-ranking procedure was applied to the highest ranked poses to increase docking accuracy. The MolDock scoring function (MolDock Score) used by MVD is derived from the piecewise linear potential scoring functions originally proposed by Gehlhaar and co-authors (Gehlhaar *et al.*, 1995) and later extended by Yang and co-authors. (Yang *et al.*, 2004). The MolDock scoring function further improves these scoring functions with a new hydrogen bonding term and new charge schemes. The docking scoring function, E_{score}, is defined by the following energy terms:

$$E_{score} = E_{inter} + E_{intra}$$

where E_{inter} is the ligand-protein interaction energy:

$$E_{inter} = \sum_{i \epsilon ligand} \sum_{j \epsilon protein} \left[E_{PLP}(r_{ij}) + 332.0 \frac{q_i q_j}{4 r_{ij}^2} \right]$$

The summation runs over all heavy atoms in the ligand and all heavy atoms in the protein, including any cofactor atoms and water molecule atoms that might be present. In the benchmarks, all water molecules were removed before docking. The second term describes the electrostatic interactions between charged atoms. The numerical value of 332.0 fixes the units of the electrostatic energy to kcal/mol. For distances less than 2.0 Å, the electrostatic energy is cut off at the level corresponding to a distance of 2.0 Å to ensure that no energy contribution can be higher than the clash penalty. Although the electrostatic energy contribution has the theoretically predicted magnitude the other energy terms are empirically motivated, and the total energy does not necessarily correlate with the true binding affinity.

Affinity grid resolution was set to 0.3 Å. Ligand evaluations were based on internal energy of binding, internal H-bonds formation, Sp^2-Sp^2 (trigonal planar electron domain geometry) torsion angles. Binding site was previously detected by CAVER module and further measured for cavity volume (282.1 $Å^3$), surface (681.0 $Å^2$) and radius (15.0 Å). Customized search algorithm was set to the MolDock SE (simplex evolution). Number of runs was 10. Parameter settings were set to 1500 iterations, 50 population sizes, 100.0 kcal/mol of energy threshold for pose generation, 300 simplex evolution steps and 1.0 neighbour distance factor. All dockings were performed at 1.0 Å RMSD threshold.

For preparing the AutoDock docking parameter file I used default settings (genetic algorithm parameters: population size = 150, number of energy evaluations = 2500000, rate of gene mutation = 0.02, rate of crossover = 0.8, maximum number of generations = 27000, number of GA runs = 10, initial dihedrals were randomly specified, elitism value was set to

1). By default, clustering of docked results was done at 0.5 Å RMSD. Prior to docking, total Kollman and Gasteiger charges were added to the protein and the ligand.

3.3.6 Construction of pharmacophore models

The PyMol LIQUID ("ligand-based quantification of interaction distributions") module was utilized to create a pharmacophore 'fuzzy' model. A conceptual idea of LIQUID is to combine multiple docking poses of a known inhibitor into a single fuzzy description of potential ligand–receptor interaction sites. This technique models a pharmacophoric features as Gaussian densities and employs trivariate models.

At the first model computation LIQUID used standard cluster radii to calculate cluster sizes at 4.0 Å RMSD for lipophilic areas and 1.9 Å RMSD for H-bond donors and acceptors (Tanrikuli *et al.*, 2007). In my experiment, I set up default cluster radius (2.0 Å) for each potential pharmacophore point type **(Figure 3.4 D)**.

3.4 Results and discussion

Genetic algorithm produced 1300 compounds (70 generations) of diverse chemical structure. 500 of top molecules were selected according to their shape similarity (0.7-0.9 functional scores). These molecules, which had pharmacophoric score less than 0.8 were excluded from the list. 227 out of 500 compounds satisfied all that criteria and were tested for Lipinski's Rule of Five using the LigandScout pharmacophore software (Wolber and Langer, 2005). Only one compound (novel hit) did not have rule violations, in contrast to IDV, which had one violation of Lipinski's rule. A useful parameter, such as topological polar surface area (TPSA), was defined for these two molecules as the surface sum over all polar atoms to evaluate parameters for the prediction of cell permeability and drug transport properties **(Table 3.1)**. This parameter shows good drug transport, even in ability to penetrate the blood-brain barrier. In this case TPSA should not exceed 600 $Å^2$ threshold. The novel hit was analyzed for Cramer rules, carcinogenicity and biodegradability. This observation also showed a strong affinity (in comparison with reference molecule) to the HIV-1 protease binding site. Docking results revealed the interaction modes for this compound with the HIV-1 protease. Novel hit adopts IDV binding mode (Asp 25, 29 and Gly 27) and also has the additional interacting residues (Gly 49-Ile 50, Asp 29-Ala 28, Asp 30). Moreover, it exceeds IDV in H-

bond formations (10 instead of 9). New compound's nitrile group is creating H-bonds with donor nitrogen of Gly 49 - Ile 50 amide group **(Figure 3.4 B, C).**

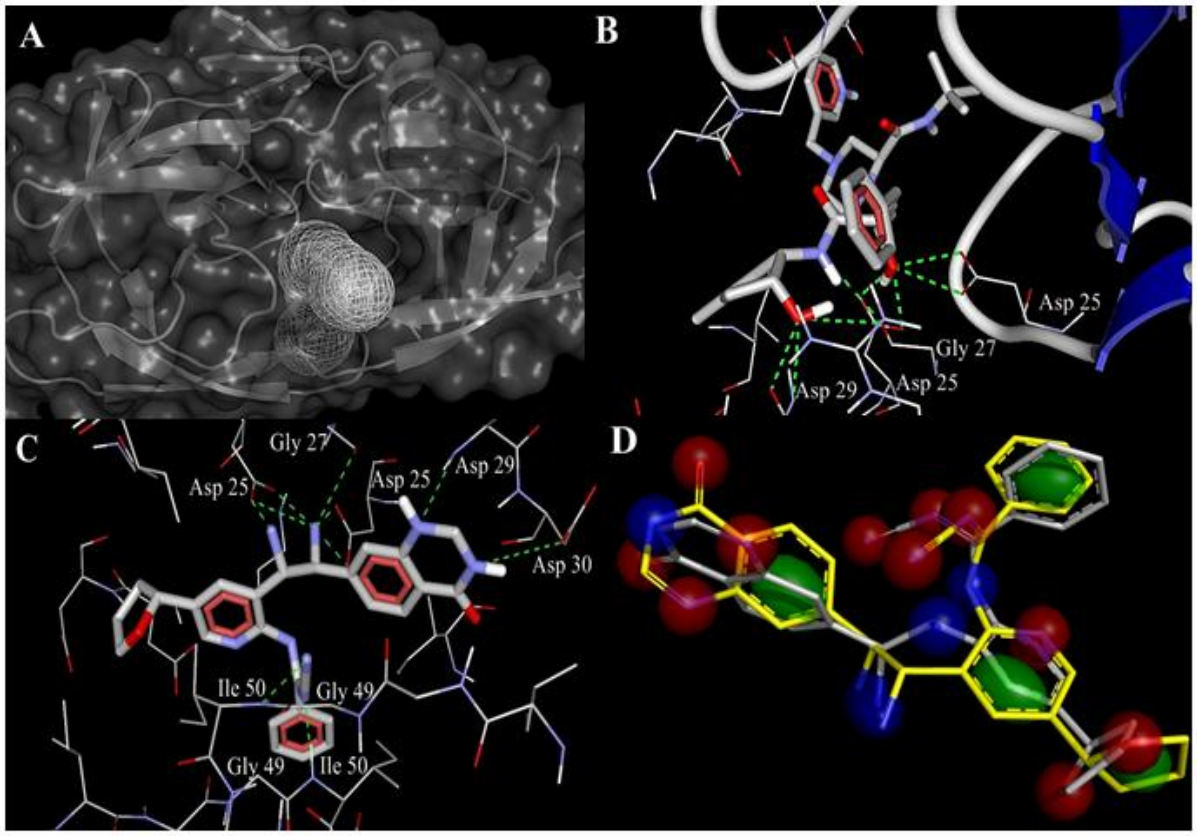


Figure 3.4: (A) HIV protease catalytic tunnel (binding site) was predicted by the PyMol CAVER module. IDV (B) and novel hit (C) interactions with the HIV-1 protease are shown. H-bonds are depicted as dashed lines. IDV – protease complex was analyzed as a crystal structure. 3D alignment and 'fuzzy' model (D) of the hit 'native' conformation (gray) together with its 'functional' conformation (yellow) at 2.0 Å RMSD. The shown potential pharmacophore points are colour-coded as follows: lipophilic areas are green, H-bond donors and acceptors are coloured in blue and red respectively. The pictures were generated using the PyMol software.

Probably this acceptor nitrile group along with the tetrahydropyridine ring might have effect on the ligand conformational shift with subsequent change in binding affinity mode. In both docking cases, the catalytic function of the Asp 25 residues is completely blocked. Subsequently, I performed the AutoDock experiments to validate the above data and revealed reduction in hydrogen bonds formation presumably due to differences in affinity grid resolution of the Molegro Virtual Docker and the AutoDock (0.3 vs 0.375 Å). However, consistent

with previous observation, all crucial amino acids obtained from Molegro docking experiments are also present in the AutoDock docking profiles **(Figure 3.5 A, B)**. The IDV spatial conformation recruits less docking energy (-15.1 kcal/mol) in comparison to novel hit and forms H-bonds with amino acid residues of chain B. In this orientation, the amide nitrogen donor together with carboxyl group (hydroxyl moiety) of the Asp 29 residue, the Asp 25 carboxyl group (hydroxyl moiety) and the Gly 27 amide oxygen acceptor make hydrogen bonds with two hydroxyl functional groups and one amide nitrogen donor of the ligand molecule.

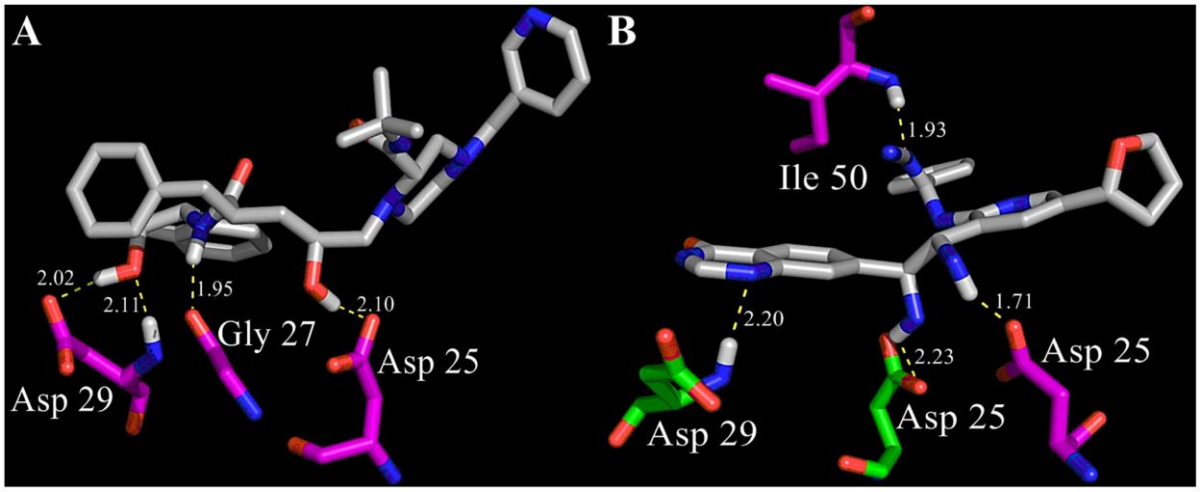


Figure 3.5: (A) Indinavir and (B) novel hit docking profiles calculated by the AutoDock software. Both ligands are coloured in gray. Chain A and B amino acids of the HIV-1 protease are painted in green and magenta respectively. The hydrogen bonds are shown for each of both molecules; see text for details. The pictures were generated using the PyMol software.

Although, novel hit conformation interacts with both chains of the HIV protease but its global minimum in docked energy (-13.8 kcal/mol) with the protease binding site is more than that of IDV, resulting in decrease of binding affinity and intermolecular force between the ligand and its receptor. The Asp 29, Ile 50 amide nitrogen donors form H-bonds with N^1 of the tetrahydropyrimidine ring and nitrogen acceptor of the aminoformonitrile residue. Carbonyl and hydroxyl moieties of both Asp 25 carboxyl groups are interacting with amine functional donor groups of novel hit. Asp 25, Asp 29, Gly 27 residues are located in the 'eye' areas, whereas Ile 50 residue is localized in the 'flap' region. The 'flap' region of the protease plays an important role in the enzyme "opening" and ligand binding to the 'flap' β-hairpins could block "opening" process. Finally, I measured the hydrogen bond length by measuring the distance between the donor and acceptor atoms. Usually, longest H-bond length is about

3.5 Å. Anything longer of this parameter would be considered a pure dipole-dipole interaction. It is well established that hydrogen bonds have a typical length around 2.5 Å. In my case, all H-bond distances were in the range from 1.71 to 2.11 Å. Interestingly, one bond in IDV and two bonds in the hit molecule were detected as ultra-short hydrogen bonds with donor and acceptor distances of less than 2.0 Å. This observation supports further that identified the novel hit compound should have a tendency to form strong hydrogen bonds to the protease catalytic channel. One drawback to IDV and novel hit is that, they both have a heterocyclic rings with complex substituent. Those features are responsible for the third class of toxicity (highly toxic). Both molecules belong to the second class of biodegradability, which means no alerts for notifying an easily degradable chemical were found and there was one alert notifying a persistent chemical the compound under investigation is declared as "persistent chemical".

Table 3.1 Comparative characteristics of IDV and novel hit

Molecular Properties	**IDV**	**Novel hit**
Atoms	92	63
MW (≤ 500)	613.8	494.6
TPSA ($Å^2$)	118	181
cLogP (≤ 5)	2.9	2.4
HBD (≤ 5)	4	4
HBA (≤ 10)	7	6
NI[1]	0	0
PI[2]	2	3
Docking results	**IDV**	**Novel hit**
MolDock score	-	-185.7
Rerank score	-	-148.5
Docking score	-	-191.6
H-bond energy*	-	-12.8
Flexible torsions	-	8
Energy minimum*	-15.1	-13.8
ADME/Tox	**IDV**	**Novel hit**
Cramer rules	Class 3	Class 3
Biodegradability	Class 2	Class 2
Carcinogenicity	no	no

[1] *negative ionizable groups*
[2] *positive ionizable groups*
[*] *kcal/mol*

3.5 Conclusions

A theoretical study was performed on a set of 1300 compounds that are structurally similar to the AIDS drug Indinavir. I generated library of diverse chemical compounds on the bases of Indinavir and screened them for possible anti-protease activity. However, mutagenic efficiency of HIV is very high in generating new viral strains and could be bypassed by creating drugs with different binding properties and strong affinity to the target protein.

The results are shown in this report, indicate that at least one compound has binding residues different from the initial hit, with improved affinity to the protease as well as its QSAR properties. Further studies on this compound will provide useful information towards the rational design of HIV-1 protease inhibitors and their effectiveness.

Chapter 4

4.1 Overview

Using classical all-atom molecular dynamics simulation, I investigated the molecular dynamics of palmitoyloleoylphosphatidylcholine and palmitoyloleoylphosphatidylethanolamine membrane bilayers enforced by a single-wall carbon nanotube. I postulated that an insertion of a single-wall carbon nanotube in the center of lipid membrane 'strengthens' ambient lipids and prevents the whole system from further destabilization by high temperatures. I implemented root mean square deviation and root mean square fluctuation analyses of simulated structures from their initial states in order to emphasize the molecular dynamics behavior of these structures during 1000 ps simulation time at different temperatures. The data suggest that an intercalated carbon nanotube restrains the conformational freedom of adjacent lipids and hence has an impact on the membrane stabilization dynamics. On the other hand, different lipid membranes may have dissimilarities due to the differing abilities to create a bridge formation between the adherent lipid molecules. The results derived from this work may be of importance in developing stable nanosystems for construction of novel biomaterials and delivery of various biomolecules in the fields of biosensors, biomaterials, and biophysics.

4.2. The simulation and its goals

The importance of nanotechnology for biotechnological applications is frequently discussed in the scientific community as a powerful tool for the development of nanostructured materials and (Zhou *et al.*, 2007), ranging from novel nanoelectrics to molecular assemblies, to nanobiocomposites, tissue engineering and biomedicine. The physico-chemical properties of some of these nanomaterials determine their biocompatibility via supporting and stabilizing different biological systems such as lipid membrane bilayers by means of their integration into functional hybrids (Zhou *et al.*, 2007).

The most basic property of lipid bilayers in biology is to act as a barrier between the inside and outside of a cell, and this depends on the continuity, molecular composition, stabil-

ity, etc of the bilayer structure. 'Black' lipid membranes may self-organize from bilayer-forming phospholipids, which are quite stable at room temperature (Stern *et al.,* 1992). Conversely, high temperatures, pressure electrolyte imbalance and other extreme physiological and environmental conditions lead to irreversible phase separation of lipid components within the membrane (Quinn, 1989).

Moreover, the unbalanced electric stress from the external electric field created by an ionic current across a lipid membrane bilayer gives rise to a destabilizing surface energy enhancing undulations (Sens and Isambert, 2002) and multiple pore formation as a result of electroporation (Tarek, 2005).

It was recently shown that a palmitoyloleoylphosphatidylcholine (POPC) membrane bilayer is supported by a hydrophobic carbon nanotube (CNT) network, creating a mechanically strong surface and increasing structural stability (Gagner *et al.,* 2006). On the other hand the CNT is able to penetrate inside mammalian cells, and the cellular internalization of CNT occurs without the help of any external transporter (Kam *et al.,* 2005; Kam *et al.,* 2006). Alternatively, stable lipid-coated CNTs have been developed in aqueous media via noncovalent wrapping of the lipid chains onto graphitic surface of single-wall carbon nanotube (Douroumis *et al.,* 2007).

Several studies indicated that CNTs can be spontaneously inserted into lipid membrane bilayer (Vakarelski *et al.,* 2007; Bianco *et al.,* 2005; Cherukuri *et al.,* 2004; Lu *et al.,* 2004), but the stabilizing effect of intercalated carbon nanotube on the lipid membrane bilayer is not well understood. A better understanding of the atomic level interactions between the CNT and lipid membrane is also necessary in designing and developing superior drugs, cellular injections equipment and gene therapy vectors.

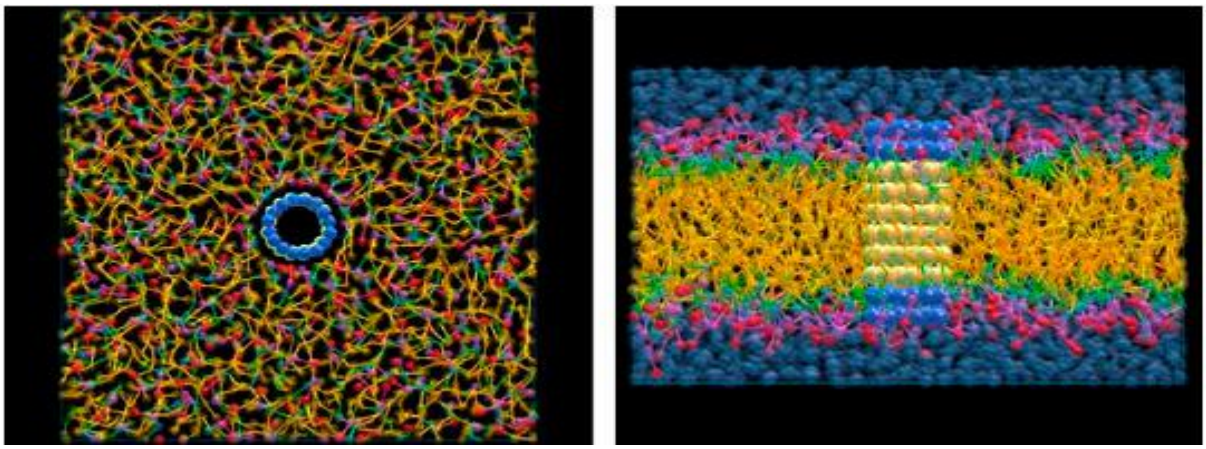


Figure 4.1: Snapshots, from the top and the side, of the membrane-CNT unit cell consisting of 256 DMPC lipids, 2560 coarse grain water sites representing 7680 water molecules, and one 10-ring narrow intercalated single-wall carbon nanotube. The six inner hydrophobic CNT rings are coloured white whereas the hydrophilic rings are coloured blue. The lipid tails are shown in yellow and the headgroups in red, purple, and green. The water, which is suppressed in the top view, is coloured in blue (figure modified after Nielsen et al., 2009).

Unfortunately, lipid membranes are very fragile and their molecular dynamics, as well as membrane - CNT molecular dynamics behaviour and stability, are difficult to characterize using conventional *in vitro* and *in vivo* methods. The complex nature of the interactions of the lipid molecules with one another and with the nanotube and solvent molecules considers them as existing models of hydrated lipid bilayers in solution and makes analysis of these interactions quite challenging. However, computational methods have been used in recent years to tackle this issue (Peter and Hummer, 2005; Garate *et al.,* 2009; Marrink and Berendsen, 1996; Marrink *et al.,* 1998).

Since the pioneer paper entitled 'Molecular dynamics simulation of POPC at low hydration near the liquid crystal phase transition' (Ceccarell and Marchi, 1998) was published in 1998, a series of investigations into the lipid membrane and the CNT from a dynamic point of view have been stimulated. These studies have suggested that single-wall carbon nanotubes may trigger self-assembly with lipids and detergents via a non-covalent absorption mechanism of these components onto the nanotube surface (Wallace and Sansom, 2009; Wang *et al.,* 2009; Wang *et al.,* 2009; Qiao and Ke, 2006) and may serve as templates for polymerized lipid assemblies (Thauvin *et al.,* 2008) **(Figure 4.1; 4.2)**.

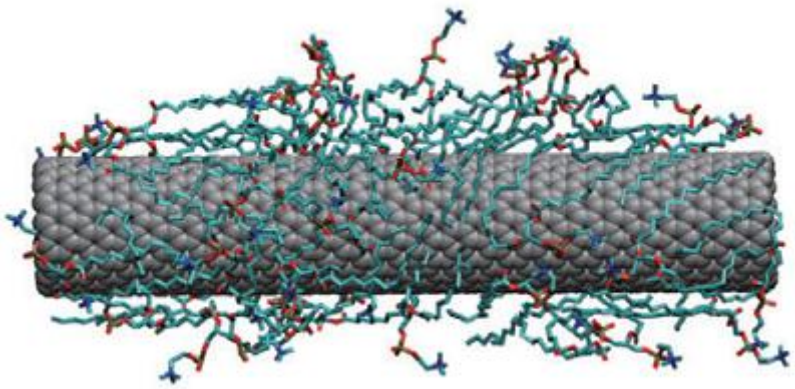


Figure 4.2: The self-assembly of DPPC and single-wall CNT for 46 DPPC molecules using GROMOS96 parameters. Front view of the simulation system configuration at 15 ns (figure modified after Wang et al., 2009).

Liu et al. showed a phenomenon of hindered lipid diffusion through the membrane bi-layer which contained a transmembrane single-wall carbon nanotube. The decrease in local diffusion coefficient of the lipid molecules was directly-proportional to the CNT diameter in the lipid membrane (Liu *et al.,* 2009). In the molecular dynamics simulations of *Nielsen et al*. the total length of CNT was varied to monitor the hydrophobic effects at the nanotube-lipid boundary, and the lipid molecules in close proximity to the CNT showed a hydrophobic mismatch resulting in change of bilayer thickness profile (Nielsen *et al.,* 2005).

In view of this, to really understand the interaction of lipid membrane bilayers with the single-wall carbon nanotube, I should consider not only the static structures concerned but also the dynamical information obtained by simulating their interactions through a molecular dynamic process.

In the present section, I demonstrate the phenomenon of the lipid membrane instability due to high temperatures, and propose the root mean square deviation (RMSD) and root mean square fluctuation (RMSF) analyses for comparison of simulated systems via the optimal superimposition of either full structures and prescribed substructural elements or defined atomic positions after first fitting to a reference frame. Finally, a least-square fit procedure is utilized and followed by computation of the average distances between the corresponding atoms in the dynamically analyzed system.

A distribution of pairwise RMSD values provides information on the similarity of members of a given ensemble when it comes to their overall structure. However, to get information on local structural flexibility, thermal stability and heterogeneity of simulated molecules, RMSF parameter has been often studied (Benson and Daggett, 2008; Priyakumar *et al.,* 2010; Pan and MacKerell, 2003).

Here, I used a high temperature (400 K) molecular dynamics simulation, because it is known that critical fluctuations of lipid membranes may occur even at 313 K, and above 343 K the interlamellar water layer thickness starts to increase non-linearly due to ‘hydration force’ (Kirchner and Cevc, 1993). I implemented classical dynamics simulation to emphasize POPC and POPE (palmitoyloleoylphosphatidylethanolamine) membrane bilayer stability with or without an intercalated single-wall carbon nanotube.

4.3 Computational methods

In the present study, I investigated classical molecular dynamics of five engineered *ab initio* systems (CNT, POPC, POPE, POPC-CNT and POPE-CNT) at different temperatures (300, 300-400 and 400 K). To model a membrane bilayer, I built a 3D rectangular matrix of the required size (20×20×42 Å) out of pre-equilibrated and pre-hydrated membrane square patches as a 2-dimensional hexagonal lattice of lipids. To make the generated lipid membrane structure closer to an actual one, some disorder was introduced into the patches: random orientation of each lipid in the membrane plane, and truncated Gaussian spread in the perpendicular direction. The membrane building tool, version 1.1 (Balabin, 2008) was used to construct the membranes in their liquid crystal lamellar phase. The lipid tails were fully extended, allowing for easy inserting of the CNT cylindrical shape into the membrane and, therefore, reducing the required equilibration time. To properly hydrate the lipid head groups, water shells were built around the lipids one by one. The distance between the layers was set to fit the actual membrane thickness, and the lattice period was set to fit the actual surface density of lipid molecules. Proper head group hydration may be critical for membrane properties and, therefore, for function of membrane proteins. To properly hydrate the lipid head groups, water shells are built around the lipids one by one. Water molecules outside the lipid dimensions and inside the hydrophobic layer (there usually are only a few) are deleted. The final patch structure is a slightly disordered lipid bilayer with the lipid head groups solvated.

The POPC (POPE) membranes had the following parameters: atoms = 1146 (1077), bonds = 1026 (962), angles = 1650 (1537), dihedrals = 2136 (1974), residues = 120 (115), waters = 114 (109), segments = 3 (5), fragments = 120 (115). To simulate the CNT, I have built relatively short (54.05 Å in length) single-wall carbon nanotube with the diameter of 6.80 Å. It consisted of 920 atoms and 1135 covalent bonds. To construct an armchair CNT **(Figure 4.3)**, I produced the unit cell of a CNT by first generating coordinates of a rectangular graphene sheet with dimensions corresponding to the particular chiral indices (*n*, *m*) according to the equation below:

$$C = na_1 + ma_2$$

The length of *C* equals the circumference of the resulting CNT. Thus, the CNT diameter *d* is given by

$$d = \frac{|C|}{\pi} = a\sqrt{(n^2 + nm + m^2}$$

Here, a = 2.49 Å is the lattice constant of the graphene honeycomb lattice. The translation vector *T* connects two equivalent carbon atoms along the CNT axis and is given by

$$T = \frac{|(2m+n)a_1 - (m+2n)a_2|}{\gcd(2m+n, m+2n)}$$

where *gcd* is the greatest common divisor function. The length of *T* gives the length of the SWCN unit cell and can be written as:

$$|T| = \frac{\sqrt{(3)d}}{dcg(2m+n, m+2n)}$$

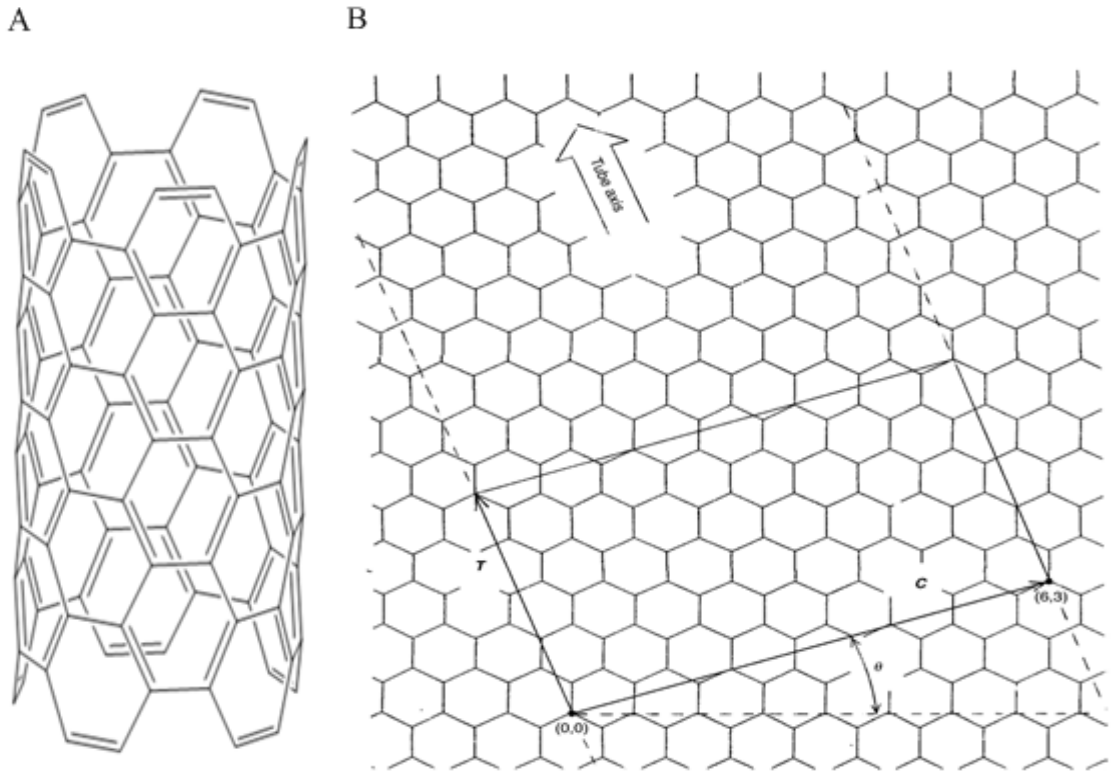


Figure 4.3: (A) Armchair carbon nanotube formed from a graphite sheet that is rolled up so that the edge is in the shape of armchairs. High symmetry armchair CNTs occur for m = n chiral indices, respectively. (B) The construction of the unit cell for a CNT (see the previous section for details) (figure modified after Dresselhaus et al., 1995).

These coordinates are then mapped onto a cylinder to generate the CNT unit cell with its central axis aligned in the z-direction. The unit cell is then replicated along the z-axis to generate a SWCN of the desired length.

The nanotube was manually aligned in a rectangular manner to the horizontal (x-axial) plane position of the lipid membrane and embedded in its center **(Figure 4.4 A-C)** by

means of the NanoEngeener-1 software (Nanorex, Inc.). To simplify the computational approach only undecorated CNT was used in the molecular dynamics simulation.

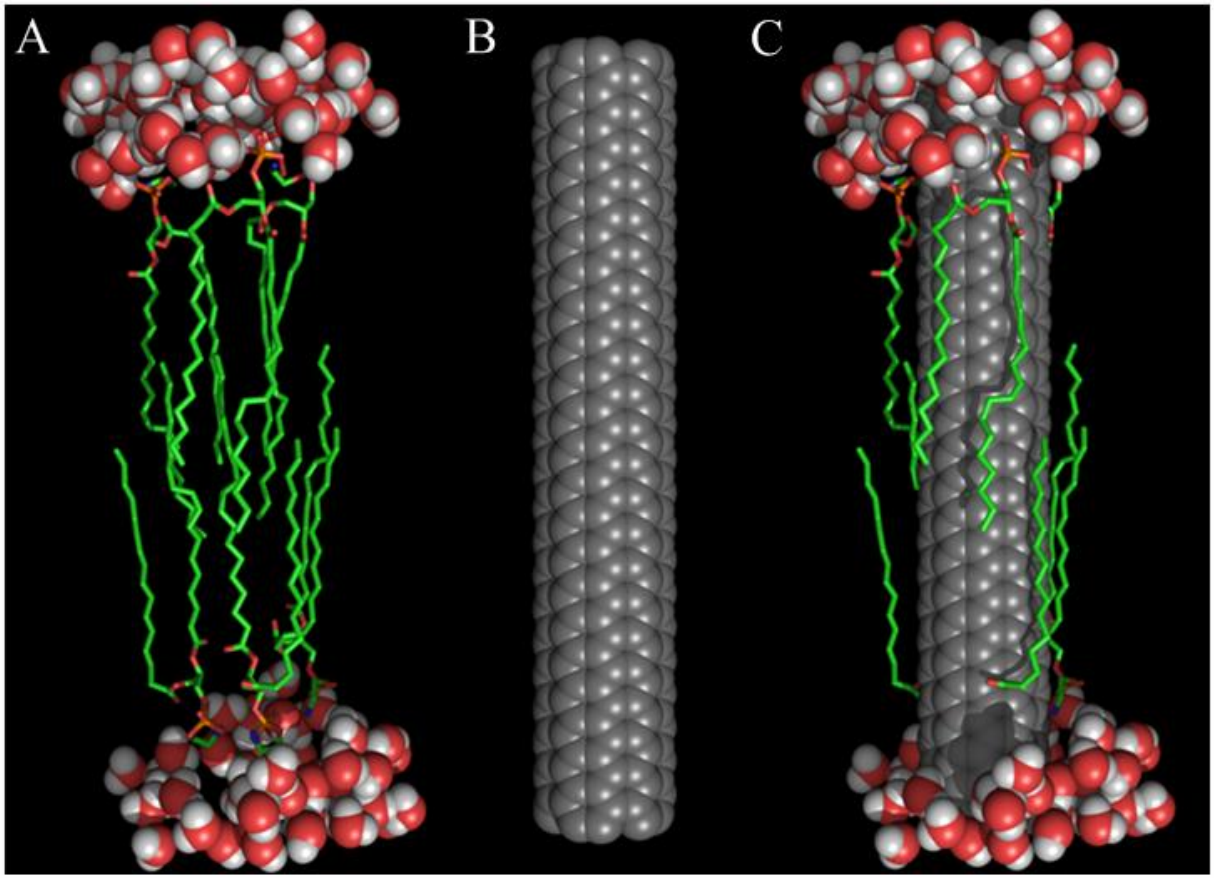


Figure 4.4: Schematic representation of the lipid membrane bilayer stabilized by a single-wall carbon nanotube: (A) Lipid membrane; (B) CNT structure; (C) Membrane-CNT complex. CNT (hydrogen atoms removed), water and lipid molecules are given in 'space - filling' and 'steak' representations, respectively. The pictures were generated using the PyMol software.

The quantitative assignment of the dynamic stability was done by measuring the RMSD of the carbon atoms from their initial positions in the system by least-square fitting the structure to the reference structure using an equation, namely:

$$RMSD_{\alpha} = \sqrt{\frac{\sum_{\alpha=1}^{N_{\alpha}} \left(\vec{r}_{\alpha}\left(t_j\right) - \langle \vec{r}_{\alpha} \rangle\right)^2}{N_{\alpha}}}$$

where N_{α} is the number of atoms whose positions are being compared, N_t is the number of time steps over which atomic positions are being compared, $\vec{r}_{\alpha}(t_j)$ is the position of atom α at

time t_j, and $\langle \vec{r}_\alpha \rangle$ is the average value of the position of atom α to which the positions $\vec{r}_\alpha(t_j)$ are being compared. Finally, it is defined as:

$$\langle \vec{r}_\alpha \rangle = \frac{1}{N_t} \sum_{j=1}^{N_t} \vec{r}_\alpha (t_j)$$

The RMSF is a commonly used value of variability across multiple structure alignment. For computational reason, it was approximated as a RMSF of the carbon atoms and calculated as a measure of the deviation between the positions of particle *i* (carbon atom) and the reference position according to the equation below:

$$RMSD_i = \sqrt{\frac{\sum_{j=1}^{N} \left(C_{j,i} - \bar{C}_i\right)^2}{N}}$$

where $C_{j,i}$ is the position of the i^{th} carbon atom of the j^{th} complex and $\bar{C}_i$ is the average position for the i^{th} carbon atom from the multiple alignment of *N* structures.

Specifically, all simulations were performed using parameterized all-atom MMFF94x (Merck Molecular Force Field), modified from MMFF94s to force conjugated nitrogens planar. Taking into account a typical range of interactions in simulated systems (10 Å), to reduce the computational cost and domination of boundary effects, water molecules were added as TIP3P three-site point charge models. Due to the bilayer thickness being ~40-45 Å, the size of periodic box in the direction perpendicular to the membrane plane should be at least 50 Å or higher. Finally, the system was consisted of 591 water molecules surrounding the lipid membrane in a periodic box of 54.05 Å edge length and 21.72 Å thickness. Partial charges for the atoms to be minimized were assigned according to the MMFF94x force field prior to the energy minimization.

All MD simulations were performed using a standard two-step approach. In the first step, the energy minimization was initially performed, the partial charges for all the atoms were calculated and the simulated systems were minimized for 100 steps using 0.05 Å root mean square gradient.

In the second step, 1000 ps molecular dynamics simulation was performed. The system was sampled and written to the database after each 1 ps passage. The 2 fs time step was used in discretizing the Nosé-Poincaré-Anderson (NPA) equation of motion.

The NPA equation is based upon a Hamiltonian which provides an advantage when attempting to generate stable trajectories even in short time scales: Hamiltonians are symplec-

tic which means they conserve phase space volume (Sturgeon and Laird, 2002, Bond *et al.*, 1999). The NPA Hamiltonian is calculated as:

$$H(q,u,s,u_s,V,u_v) = s[H_{NA}(q,u,s,u_s,V,u_v) - H_o]$$

$$H_{NA}(q,u,s,u_s,V,u_v) = \frac{u^T M^{-1} u}{2S^2V^{2/3}} + U\left(V^{\frac{1}{3}}q\right) + gkTlogs + \frac{u_s^2}{2gkTm_T} + PV + \frac{u_v^2}{2gkTm_V}$$

$$H_o = N_{NA}(q,u,s,u_s,V,u_v)_{t=0}$$

where q is the scaled-space coordinates of the atoms, u is the scaled-space momenta of the atoms (conjugate to q), r is the real-space coordinates of the atoms, M is the diagonal atomic mass matrix, m_V is the squared relaxation time of the pressure control coordinate and g is the number of degrees of freedom in the atomic system plus 1 if pressure is fixed, s is the time scale coordinate and u_s its momentum. These equations preserve the Hamiltonian and represent the uniformity in time; this means that time samples and averages correspond to "true" time. To solve these equations the symplectic, time-reversible generalized leapfrog algorithm was utilized (Sun, 1993).

The NVT ensemble was used and the constant-temperature condition (T_{const} in Kelvin) was kept during molecular dynamics calculation. The number of particles (N) and the volume of each system (V) were the same for NVT ensemble. Hydrogen bonds were constrained and treated as rigid bodies. The constant pressure of 101 kPa (for the NVT simulation), 0.2 ps temperature response (m_T) and 5.0 ps pressure response (m_P) were maintained to enforce pressure and temperature parameters in a molecular dynamics simulation. The nominal temperature (T_0) was used to assign initial velocities at time 0 (start of simulation). For 300-400 K transition temperature parameter, I did not specify a heat time to reduce a 'critical increment' adaptation and probe an extreme robustness of the system **(Figure 4.5 A)**. The calculations have been performed using the Molecular Operating Environment 2007.09 (Chemical Computing Group), the GROMACS 4.0.7 (Groningen Machine for Chemical Simulations) (Van der Spoel *et al.*, 2005, Kutzner *et al.*, 2007) and the VMD 1.8.7 (Visual Molecular Dynamics) (Humphrey *et al.*, 1996) molecular modelling software suites. The RMSD and RMSF data are presented as average values ± standard deviation.

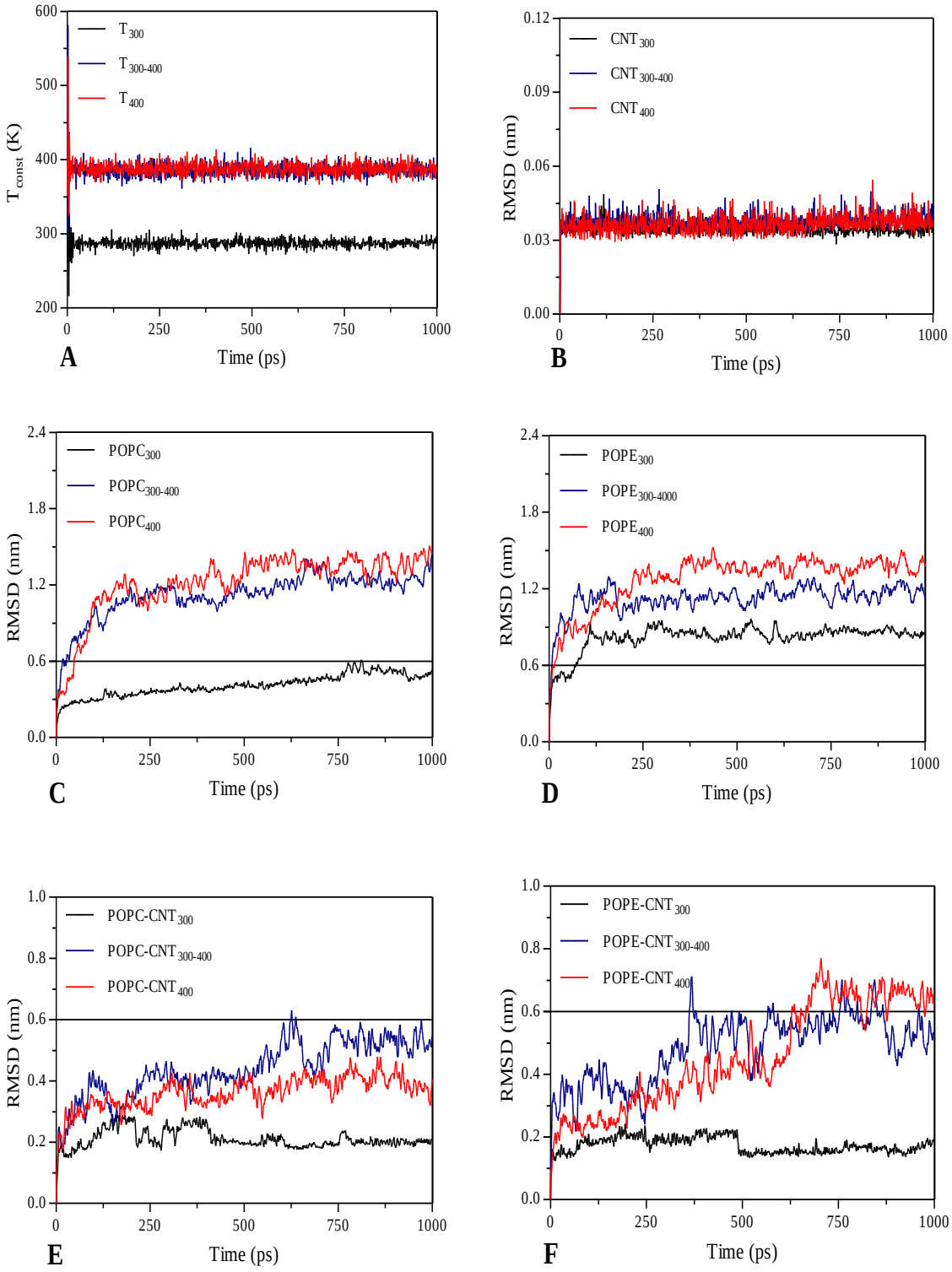
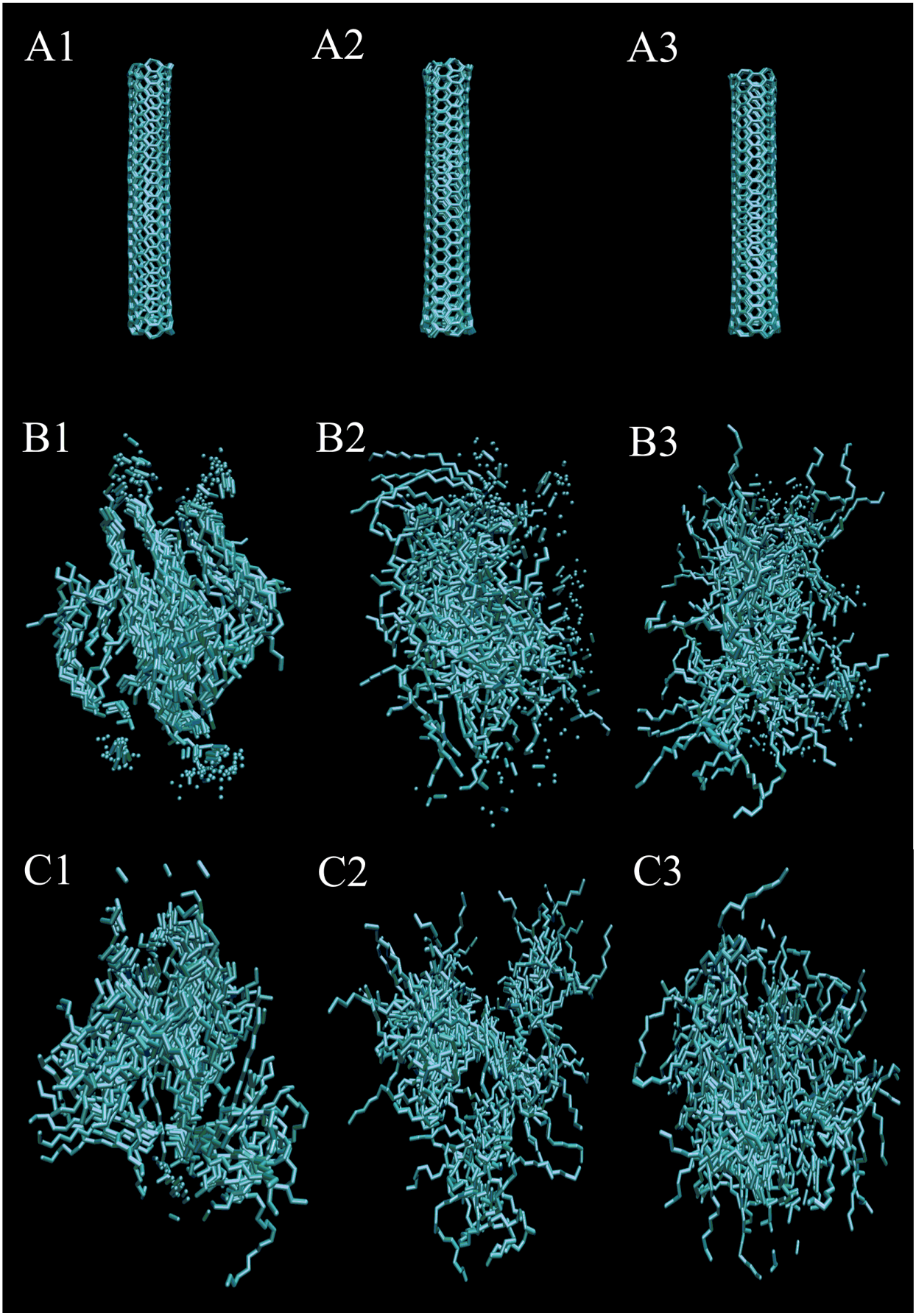


Figure 4.5: (A) Constant temperature parameters and (B-F) RMSD values are shown during 1000 ps evolution time. All RMSD values of investigated MD systems are represented with respect to their initially minimized structures at different temperature levels.

4.4. Results and discussion

Prior to membrane-nanotube complex MD simulation, I have performed molecular dynamics of the CNT and membranes (POPC, POPE) in water solution. The accurate molecular dynamics simulation of membrane-nanotube is challenging due to the polarizability of nanotube atoms and the fact that lipids exchange their positions in the bilayer plane on the time scale of a few hundred nanoseconds at 300K, the time scale of 1000 ps might not be entirely sufficient. An accurate lipid membrane simulation may therefore require extensive calculations or employment of a polarizable force-field, which may restrict simulations to short time scales, and for small lipid fragments the 1ns simulation time is adequate: only several lipid molecules are simulated.

To better characterize structural changes within MD systems, I have calculated average RMSDs for 1000 molecular frames, which specify instantaneous orientation of the molecules. The entire nanotube structure was remained intact through the simulation and predicted to be quite stable system in the aqueous environment with the RMSDs of the backbone atoms of 0.035 ± 0.002 nm at 300 K, 0.038 ± 0.003 nm at 300-400 K transition temperature and 0.037 ± 0.003 nm at 400 K **(Figure 4.5 B; Figure 4.6 A1-A3)**.


A1
A2
A3
B1
B2
B3
C1
C2
C3

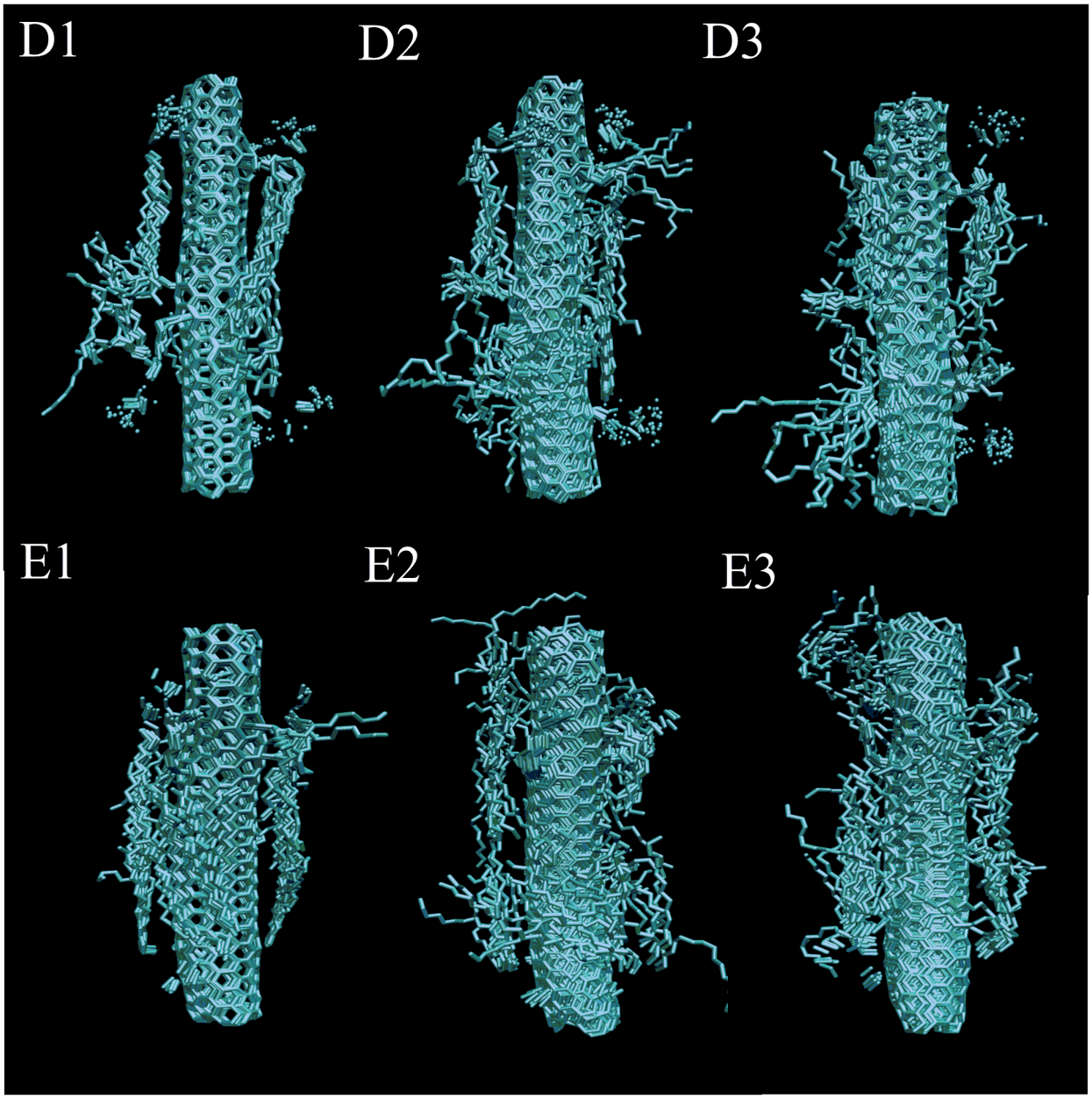


Figure 4.6: Visualization of the molecular dynamics trajectories (multiple frames) at different temperature parameters (300, 300-400 and 400 K): (A1-A3) Single-wall carbon nanotube; (B1-B3) POPC membrane; (C1-C3) POPE membrane; (D1-D3) POPC-CNT complex; (E1-E3) POPE-CNT complex. Images of every hundredth frame are shown simultaneously to make the large-scale motion of the system more apparent. Molecules are represented as carbon frameworks. The pictures were generated using the VMD software.

In the next part of the study, I have investigated the impact of molecular dynamics simulation on the stability and structural properties of POPC and POPE lipid membrane bilayers. Light-increment was observed at the $POPC_{300}$ RMSD values during the first 780 ps evolution time. The overall $POPC_{300}$ system was found to be largely below the 0.6 nm threshold with an average RMSD of 0.413 ± 0.094 nm **(Figure 4.5 C)**. Interestingly, this $POPE_{300}$ RMSD level (0.846 ± 0.106 nm) was exceeded two times in comparison with $POPC_{300}$ system, which means that the POPE membrane is prone to be less dynamically stable. In the POPC membrane this could be as a result of the ion bridge formation between the adherent molecules in the amphiphilic lipid molecule due to negatively charged phosphatidylcholine components.

The RMSD perturbation levels for POPC and especially for POPE membranes at 300K are quite high. This observation could be explained by the lamellar liquid crystal phase transition instability of the lipids, usually occurring in more advanced stage at high temperatures (Crowe *et al.,* 1988; Tsvetkov *et al.,* 1989; Koster *et al.,* 1994).

Apparently, at high hydration state, membrane hydrophilic surfaces in the water experience a strong repulsion coursed by 'hydration force', which increases approximately exponentially and leads to lipid membrane instability and destruction **(Figure 4.6 B1-B3; C1-C3)**.

More compact and less protruded POPC bilayer interface may somehow contribute to a more stable state of the membrane by preventing lipid desorption and disintegration. Probably, some discrepancies in membrane molecular dynamics may be also due to the different conformations of POPC and POPE molecules in the water environment. Further analysis of perturbations of membrane trajectories showed higher RMSDs at 300-400 K transition and at 400 K temperature **(Figure 4.5 C, D)**. In this case, the $POPC_{300\text{-}400}$ and the $POPC_{400}$ systems were characterized by average RMSD values of 1.094 ± 0.174 nm, 1.238 ± 0.245 nm and 1.147 ± 0.109 nm, 1.300 ± 0.202 nm for POPE membrane respectively, which means that all RMSDs were more or less at the same levels. These results also imply that, the lipid membrane high temperature dynamic behaviour is more affected and correlates over similar range with time.

In the last part of this investigation, I studied the dynamic stability of a membrane-CNT system through the nanotube intercalation into hydrated lipid bilayer. A RMSD levels in the membrane-CNT system were highly reduced in comparison with the previous results **(Figure 4.5 E, F)**. In the case of the POPC membrane, the POPC-CNT_{300} system yielded a RMSD value of 0.191 ± 0.035 nm, while the POPC-CNT_{300} resulted in a RMSD value of

0.175 ± 0.026 nm. The decrement was detected at RMSD values after 400 ps of simulation for the POPC-CNT_{300} and after 487 ps for POPE-CNT_{300}. This indicates a decrease in the mobility and the flexibility of lipids in the overall structure.

The high temperature RMSD levels came up with a number of values such as 0.413 ± 0.087 nm for the POPC-$CNT_{300\text{-}400}$, 0.349 ± 0.051 nm for the POPC-CNT_{400}, 0.484 ± 0.104 nm for the POPE-$CNT_{300\text{-}400}$ and 0.452 ± 0.163 nm for the POPE-CNT_{400} system **(Figure 4.7)**.

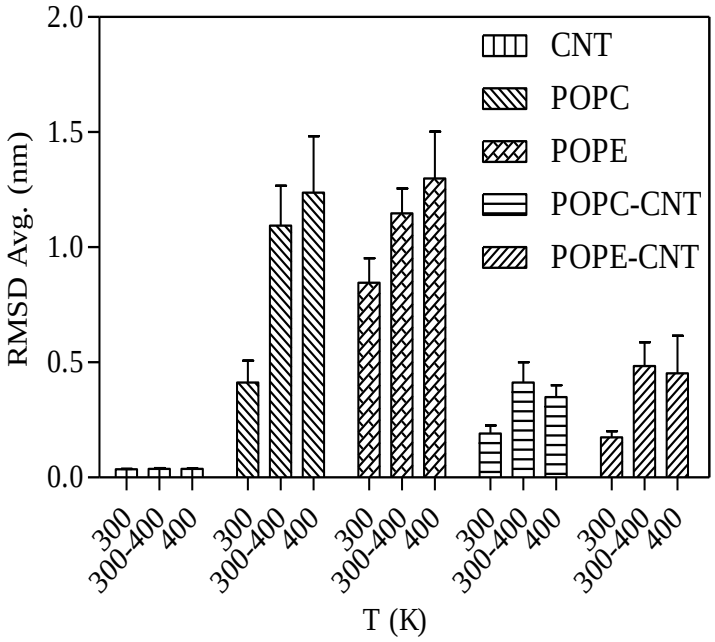


Figure 4.7: Comparative characteristics of the average root mean square deviation (RMSD) values of different simulated structures at different temperature parameters (300, 300-400 and 400 K).

Furthermore, the POPC-$CNT_{300\text{-}400}$ and the POPC-CNT_{400} RMSD trajectories notably did not exceed 0.6 nm threshold during 1000 ps molecular dynamics, whereas POPE-CNT overran this level after 670 ps of simulation at 400 K: hence the membrane-nanotube RMSD parameters were more affected at 300-400 K transition temperature (‘critical increment’) than at 400 K. It is interesting to note that the POPE-CNT_{400} RMSDs overcame the POPE-$CNT_{300\text{-}400}$ numbers after 630 ps presumably due to ‘critical increment adaptation’. The RMSD differences may be attributed to a stable environment in the membrane-nanotube system. However, the question will arise about an effect of the nanotube RMSDs in the membrane-nanotube system: Are the RMSD levels in the POPC-CNT and POPE-CNT simulations

dominate by the RMSDs from the nanotube, resulting in very low RMSD values? To answer this I performed RMSF analysis of carbon atoms for my simulated structures **(Figure 4.8 A-E)**.

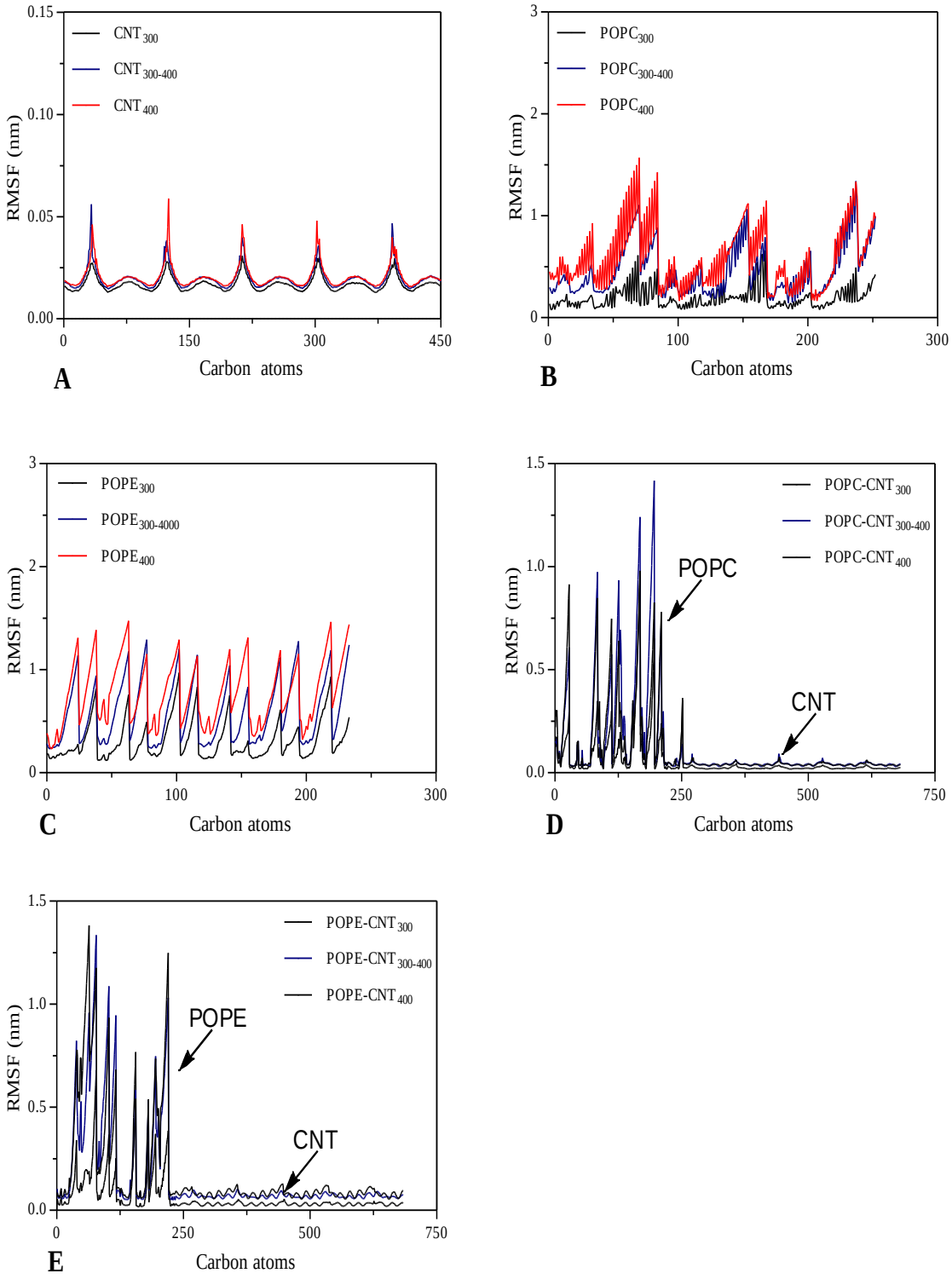
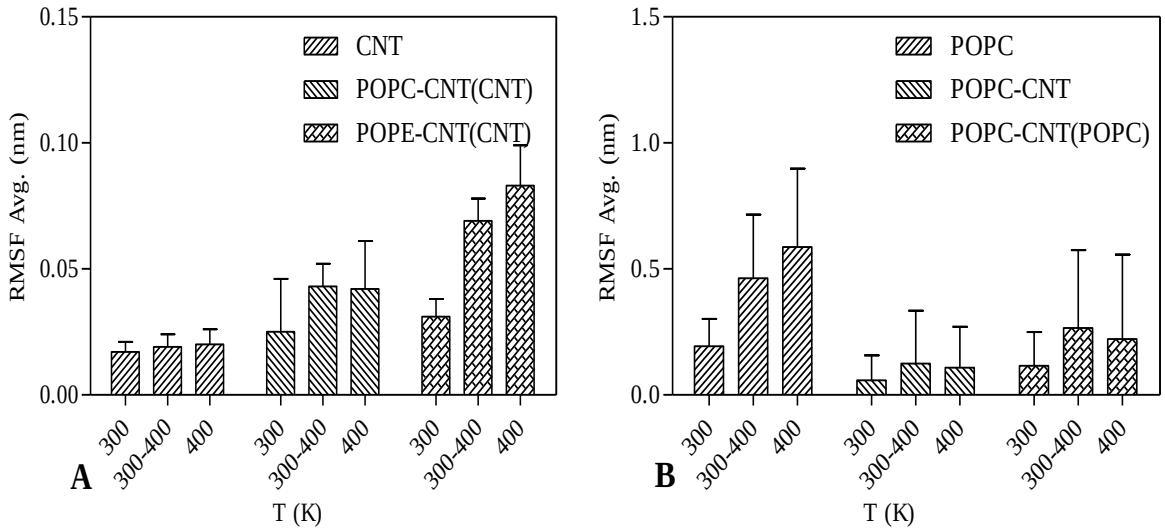


Figure 4.8: Root mean square fluctuations (RMSF) of carbon atoms at different temperature parameters (300, 300-400 and 400 K) are represented for: (A) Single-wall carbon nanotube;

(B) POPC membrane; (C) POPE membrane; (D) POPC-CNT complex; (E) POPE-CNT complex during molecular dynamics simulation. The periodic pattern shows the position of carbon atoms in CNT structure as sharp peaks interspaced by low fluctuating atoms. The calculated RMS fluctuations (fluctuations of the water molecules are not shown) show that amplitude is minimal at the ends of the nanotube. The peaks of increased flexibility are represented in the nanotube 'body' due to the low binding frequency motion of the CNT. Peaks at identical position relate to the corresponding atoms in different models. Atom numbering is from one end of the CNT to another.

In this analysis I deliberately focused on the POPC and POPE membranes, avoiding the intercalated CNT, which exhibits very low RMSF (0.017 ± 0.004 nm (300K), 0.019 ± 0.005 nm (300-400K), and 0.02 ± 0.006 nm (400K)) **(Figure 4.9 A)**. From **Figure 4.9 (B, C)** it is clear that, regardless of the nature of the lipid bilayers, for the membrane-CNT system, the lipid portion shows lower RMSF levels (0.116 ± 0.144 nm (300K), 0.266 ± 0.299 nm (300-400K), 0.222 ± 0.226 nm (400K) for POPC-CNT (POPC), 0.131 ± 0.133 nm (300K), 0.353 ± 0.308 (300-400K), 0.380 ± 0.335 (400K) for POPE-CNT (POPE)) than its counterpart at the POPC and POPE systems (0.193 ± 0.109 nm (300K), 0.463 ± 0.253 nm (300-400K), 0.587± 0.311 nm (400K) for POPC, 0.309 ±0.191 nm (300K), 0.586 ± 0.299 nm (300-400K), 0.775 ± 0.308 nm (400K) for POPE). Amphiphilic lipid molecules around the CNT may be seen as modes placed in the confined space, which imposes specific constrains upon the whole system. These constrains may prevent the system from further destabilization caused by the high temperature during conventional molecular dynamics simulation.



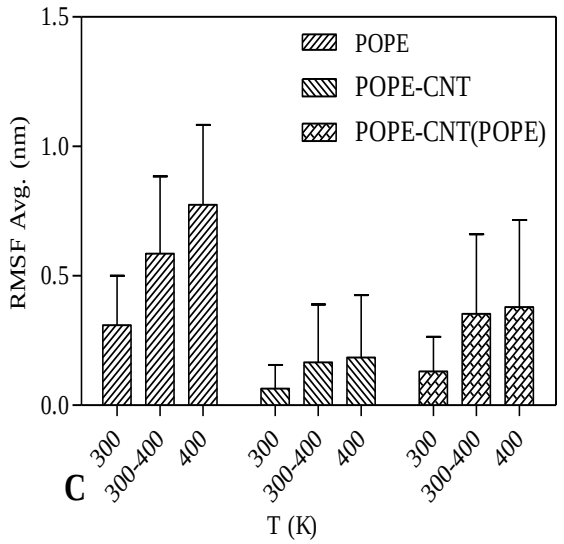


Figure 4.9: Comparative characteristics of the average root mean square fluctuation (RMSF) values of different simulated structures and substructures at different temperature parameters (300, 300-400 and 400 K): (A) RMSF average values of 'native' CNT system and the CNT substructures from different membrane-CNT systems; (B) RMSF average values of 'native' POPC, POPC-CNT systems and POPC substructure; (C) RMSF average values of 'native' POPE, POPE-CNT systems and POPE substructure. All substructural average RMSFs were calculated with respect to initial RMSF values of represented substructures, extracted from the corresponding dynamically simulated systems.

The membrane-nanotube interaction pattern shown in **Figure 4.6 (D1-D3; E1-E3**) is attributed mainly to the hydrophobic and Van der Waals interactions between the CNT and the lipid tails of the membrane bilayer. Specifically, in the 300K simulation process, when the hydrophobic lipid tails were more readily proned to adsorb on the CNT surface and the motion of the lipid tails and head-groups in the water was retarded blocking distribution of the water molecules to the CNT surface ('cylindrical encapsulation effect'). Hence, this would leave large tube surface areas unexposed to the water if, the lipid tails are generally aligned with the nanotube axis to maximize their contact with its hydrophobic surface. On the contrary, high temperatures (300-400K, 400K) create disturbances in the 'cylindrical encapsulation', which promotes the accessibility of water molecules to the CNT surface, and this situation might be confirmed by the striated formation of lysophospholipids upon the CNT, and observed in the transmission electron microscopy experiments *in vacuo* (Richard *et al.*, 2003; Wu *et al.*, 2006).

Finally, the results suggest that CNT intercalation into lipid membrane bilayer has a significant impact on the stability and the dynamic properties of the investigated systems.

4.5 Conclusions

Single-wall carbon nanotubes are a novel class of nanomaterials for the stabilization, transport and modification of biological systems into living cells. Their unique properties ensure their biocompatibility, as well as high propensity to cross and stabilize lipid membrane bilayers. While some fundamental problems of carbon nanotubes have been solved, others remain an open question. Chemical and physical interaction between CNTs and different biomolecules, such as membrane lipids, is a growing field of interest in the application of CNTs to the construction of novel biomaterials.

The complex structure of lipid membrane and its fluidity makes manipulations on a molecular and cellular level extremely challenging. However, atomistic simulations have been used with great success to understand lipid membrane-CNT interactions. To optimize these interaction processes, a realistic model of the CNT, lipid membrane and surrounding environment is needed. The present study addresses this issue using such modelling in all-atom molecular dynamics simulation.

My results show that CNT intercalation into the lipid membrane may elicit deformations in the structural organization of planar membrane architecture and hence optimize a dynamic stability of explored palmitoyloleoylphosphatidylcholine and palmitoyloleoylphosphatidylethanolamine membranes. Thus, the nanotube insertion is not affected much by high temperature dynamics, and, in this case, may stabilize the whole system.

By expanding the structural size of the simulated systems and the length of the molecular dynamics runs it will be possible to carry out more accurate and detailed studies concerning the structural changes occurring as a function of temperature, something which was not been possible in this investigation because of the high computational demand. On the other hand, a coarse-grained model of the lipid membranes has the computational power and efficiency for large-scale molecular dynamics simulation but lacks an atomistic precision. This new lipid membrane-nanotube model system has many potential applications and also can serve as a generic platform for bringing mobile lipid-bound molecules and nanotubes in close proximity for studies of both electrical and mechanical perturbations at cell membranes.

Concluding discussion

There are some viruses such as lentiviruses (HIV) that can enter the nucleus to integrate their viral genomes into the host-cell chromosome. But the mechanisms of their nuclear import and integration remain largely unknown. I tried to investigate the possible retroviral pathways into a nucleus in terms of preintegration complex (PIC) formation via HIV-1 integrase and transportin-SR2 interaction (Shityakov *et al*., 2010). However, experimental studies proved the importance of central polypurine tract (cPPT) for retroviral nuclear import (Christ *et al*., 2008). I showed that it is not possible to successfully exchange the HIV-1 cPPT-CTS for the PFV cPPT without loss of function (Shityakov and Rethwilm, unpublished data).

My own data indicates a complex and intrinsic mechanism for retroviral reverse transcription, as well as for preintegration complex formation, hence these processes cannot be duplicated in closely related viral species. Suffice it to say, there are still a lot of questions to be answered:

1) Are there some other cellular factors which could assist the PIC formation?

2) What are the exact mechanisms of their interactions with HIV-1 integrase or other retroviral proteins?

3) Is the TR-SR2 effecting/changing the spatial conformation of the HIV-1 intergase?

4) The exact multimeric structure of the HIV-1 integrase is still to be determined.

5) What is the role of nuclear pore proteins in the TR-SR2 and the HIV-1 IN interaction?

6) What is the role of foamy viral nuclear localization signals (NLSs)?

7) What is the structural organization of flap and gap elements in retroviral genomes?

8) Are there some cellular signal molecules/proteins that could bind to flap element to mediate and enhance a nuclear import?

These and many more questions regarding to the retroviral nuclear import must be answered by utilizing modern experimental and theoretical techniques. Enormous efforts have been made in the search for drugs to treat HIV infection for over decades now. After the crystal structure of the HIV-encoded protease enzyme had been elucidated (Perryman *et al*., 2004), computer-aided drug design played a pivotal role in the development of new compounds that inhibit this enzyme that is responsible for HIV maturation and infectivity. Promising representatives of these compounds have recently found their way to patients.

Protease inhibitors show a powerful sustained suppression of HIV-1 replication, especially when used in combination therapy regimens (Vacca *et al*., 1994). However, these drugs

are becoming less effective to more resistant HIV strains due to multiple mutations in the retroviral proteases (Kempf *et al*., 2001).

Computational techniques provide other options for understanding chemical systems, which yield information that is difficult, if not nearly impossible, to obtain in laboratory analysis.

I implemented computational drug design to synthesize the compound *in silico* using pharmacophoric - shape similarity scoring function and Indinavir chemical structure as a template. This virtual compound had showed the promising potential in docking studies and toxicity analysis to become a hit (Shityakov and Dandekar, 2009). Further studies must be performed on this compound as well as chemical synthesis, to measure its properties using:

1) Biochemical assays
2) Cell-based assays
3) Animal testing
4) Human clinical trials

As an important part, a detailed atomistic model was developed to investigate the interactions between single-wall carbon nanotube and different lipid membranes (Wang *et al*., 2009; Nielsen *et al*., 2009). The parameters such as CNT length, diameter, orientation, system fluctuations and deviations associated with membrane-CNT interaction in water solution were studied by simulations.

The root mean square deviation and fluctuation analyses show a significant stabilizing effect of intercalated songle-wall carbon nanotube upon palmitoyloleoylphosphatidylcholine and palmitoyloleoylphosphatidylethanolamine membrane bilayers (Shityakov and Dandekar, 2011). Carbon nanotube-lipid interactions in liquid medium have applications in a wide range of industrial processes and create challenging problems. These are some aspects that must be elucidated:

1) What is the effect of a multi-wall CNT upon a lipid membrane bilayer?
2) What is the effect of different solvents (DMSO, ethanol, methanol, etc) on an intercalated single- and multi-wall carbone nanotube and ambient lipids?
3) Application and determination of different methods/parameters such as: system's free energy, potential of mean force, coarse-grain modelling, and radial distribution functions.
4) An interesting extension of the atomistic model would be the investigation of multi-walled nanotubes interactions in different solvents/surfactants/polymers systems.

5) From the high CPU computational applications point of view, the coupling of lipid membrane coarse-grain model with a carbon nanotube all-atoms model would be an interesting extension.
6) For many important biomedical applications, atomistic model can be extended to predict the heat and mass transport phenomenon through intercalated CNTs.
7) Chemical synthesis of lipid-CNT nanobiocomposites.

Conclusion: Molecular modelling and simulation are rather new research and development methodologies, unfamiliar to many bench scientists and still considered not trustworthy. However, the unifying theme of molecular modelling and simulation methods is largely in the scope of biomedicine, which has broadened to include different approaches used to study the molecular, cellular, developmental, structural, functional and medical aspects.

By and large, one might think of computer simulations as docking studies, QSAR/ADME, molecular dynamics simulations, etc, at least as not reliable ones. However, molecular modelling credibility has steadily improved over time and in some cases has reached a level that matches the accuracy of good experimental methods.

In the case of proteins, when they are structurally known, computational modelling has evolved in a way to describe their structure and function with atomic resolution and rather straightforward, though it is not necessarily an easy task to modify proteins computationally towards a desired technical demanding.

In this regard, modified molecular modelling techniques were implemented by me to successfully solve a different biological phenomena and biochemical tasks (retroviral proteins interaction, lead optimization and nanobiocomposite formation) belonging to interrelated subareas of biology and medicine. Using pharmacophore-similarity scoring function in molecular docking process gives me a powerful tool to increase a veracity of computer simulations. Some modifications were made in the course of nanobiocomposite construction and molecular dynamics simulation e.g. periodic boundary conditions were optimized for appropriate simulated systems and Nosé-Poincaré-Anderson (NPA) equation of motion was implemented in the algorithm.

Summary

Molecular modelling and simulation are powerful methods in providing important information on different biological systems to elucidate their structural and functional properties, which cannot be determined in experiment. These methods are applied to analyse versatile biological systems: lipid membrane bilayers stabilized by an intercalated single wall carbon nanotube and retroviral proteins such as HIV protease and integrase.

HIV-1 integrase has nuclear localization signals (NLS) which play a crucial role in nuclear import of viral preintegration complex (PIC). However, the detailed mechanisms of PIC formation and its nuclear transport are not known. Previously it was shown that NLSs bind to the cell transport machinery e.g. proteins of nuclear pore complex such as transportins. I investigated the interaction of this viral protein HIV-1 integrase with proteins of the nuclear pore complex such as transportin-SR2 (Shityakov *et al.*, 2010). I showed that the transportin-SR2 in nuclear import is required due to its interaction with the HIV-1 integrase. I analyzed key domain interaction, and hydrogen bond formation in transportin-SR2. These results were discussed in comparison to other retroviral species such as foamy viruses to better understand this specific and efficient retroviral trafficking route.

The retroviral nuclear import was next analyzed in experiments regarding the retroviral ability to infect nondividing cells. To accomplish the gene transfer task successfully, retroviruses must efficiently transduce different cell cultures at different phases of cell cycle. However, promising and safe foamy viral vectors used for gene transfer are unable to efficiently infect quiescent cells. This drawback was due to their inability to create a preintegration complex (PIC) for nuclear import of retroviral DNA. On the contrary, the lentiviral vectors are not dependant on cell cycle. In the course of reverse transcription the polypurine tract (PPT) is believed to be crucial for PIC formation.

In this thesis, I compared the transduction frequencies of PPT modified FV vectors with lentiviral vectors in nondividing and dividing alveolar basal epithelial cells from human adenocarcinoma (A549) by using molecular cloning, transfection and transduction techniques and several other methods. In contrast to lentiviral vectors, FV vectors were not able to efficiently transduce nondividing cell (Shityakov and Rethwilm, unpublished data). Despite the findings, which support the use of FV vectors as a safe and efficient alternative to lentiviral vectors, major limitation in terms of foamy-based retroviral vector gene transfer in quiescent cells still remains.

Many attempts have been made recently to search for the potential molecules as possible drug candidates to treat HIV infection for over decades now. These molecules can be retrieved from chemical libraries or can be designed on a computer screen and then synthesized in a laboratory. Most notably, one could use the computerized structure as a reference to determine the types of molecules that might block the enzyme. Such structure-based drug design strategies have the potential to save off years and millions of dollars compared to a more traditional trial-and-error drug development process.

After the crystal structure of the HIV-encoded protease enzyme had been elucidated, computer-aided drug design played a pivotal role in the development of new compounds that inhibit this enzyme which is responsible for HIV maturation and infectivity. Promising representatives of these compounds have recently found their way to patients. Protease inhibitors show a powerful sustained suppression of HIV-1 replication, especially when used in combination therapy regimens. However, these drugs are becoming less effective to more resistant HIV strains due to multiple mutations in the retroviral proteases.

In computational drug design I used molecular modelling methods such as lead expansion algorithm (Tripos®) to create a virtual library of compounds with different binding affinities to protease binding site. In addition, I heavily applied computer assisted combinatorial chemistry approaches to design and optimize virtual libraries of protease inhibitors and performed *in silico* screening and pharmacophore-similarity scoring of these drug candidates. Further computational analyses revealed one unique compound with different protease binding ability from the initial hit and its role for possible new class of protease inhibitors is discussed (Shityakov and Dandekar, 2009).

A number of atomistic models were developed to elucidate the nanotube behaviour in lipid bilayers. However, none of them provided useful information for CNT effect upon the lipid membrane bilayer for implementing all-atom models that will allow us to calculate the deviations of lipid molecules from CNT with atomistic precision. Unfortunately, the direct experimental investigation of nanotube behaviour in lipid bilayer remains quite a tricky problem opening the door before the molecular simulation techniques. In this regard, more detailed multi-scale simulations are needed to clearly understand the stabilization characteristics of CNTs in hydrophobic environment.

The phenomenon of an intercalated single-wall carbon nanotube in the center of lipid membrane was extensively studied and analyzed. The root mean square deviation and root

mean square fluctuation functions were calculated in order to measure stability of lipid membranes.

The results indicated that an intercalated carbon nanotube restrains the conformational freedom of adjacent lipids and hence has an impact on the membrane stabilization dynamics (Shityakov and Dandekar, 2011). On the other hand, different lipid membranes may have dissimilarities due to the differing abilities to create a bridge formation between the adherent lipid molecules. The results derived from this thesis will help to develop stable nanobiocomposites for construction of novel biomaterials and delivery of various biomolecules for medicine and biology.

Zusammenfassung

Molekulare Modellierung und Simulationen sind leistungsstarke Methoden, um wichtige Informationen von verschiedenen biologischen Systemen, welche nicht durch Experimente erschlossen werden können, darzustellen, und deren strukturelle und funktionelle Eigenschaften aufzuklären.
Diese Arbeit untersucht in Simulationen Interaktionen viraler Proteinen sowie von Kohlenstoffenanoröhren mit Membranen und Proteinen.

Die HIV-1 Integrase besitzt Kernlokalisierungssignale („nuclear localization signals [NLS]“), welche eine entscheidende Rolle beim Import des viralen Präintegrationskomplexes („preintegration complex [PIC]“) in den Zellkern spielen. Die Ausbildung des PIC und sein Import in den Zellkern sind im Detail noch nicht bekannt. Es wurde bereits gezeigt, dass die NLS an Moleküle des Zelltransportsystems binden, wie z.B. an Transportinkernporen. Im Rahmen meiner Arbeit untersuchte ich die Interaktionen der viralen HIV-1 Integrase mit Proteinen der Kernporen wie dem Transportin-SR2 Protein (Shityakov *et al.*, 2010). Hierbei wurden die möglichen Interaktionen des Transportin-SR2 Protein mit der HIV-1-Integrase und die Bedeutung dieser Interaktionen mit dem Import in den Kern aufgezeigt. Zudem wurden die Interaktionen der Schlüsseldomänen und die Ausbildung von Wasserstoffbrückenbindungen im dem Transportin-SR2 Protein untersucht. Die Ergebnisse wurden mit Proteinkomplexen andere retroviralen Spezies, wie z.B. dem humanen Spumaretrovirus („human foamy virus [HFV]“), verglichen, um diesen spezifischen und sehr effizienten retroviralen Transportweg in die Wirtszelle zu entschlüsseln.

Der experimentelle Teil dieser Arbeit beschäftigte sich damit, den retroviralen Kernimport zu untersuchen, um die Fähigkeit des Retrovirus, nicht teilende Zellen zu infizieren, besser zu verstehen verstanden wird. Um dies zu bewerkstelligen, müssen Retroviren Zellkulturen in verschiedenen Stadien des Zellzyklus effizient transduzieren. Vielversprechende und sichere- HFV- Vektoren, welche in der Gentherapie eingesetzt werden könnten, sind nicht in der Lage, diese Effizienz bei ruhenden Zellen zu gewährleisten. Dies rührte daher, dass diese nicht in der Lage waren, einen PIC für den Transport der retroviralen DNA auszubilden. Lentivirale Vektoren sind dagegen nicht auf einen bestimmten Zellzyklus angewiesen. Für die reverse Transkription ist der Polypurinteil („polypurine tract [PPT]“) essentiell für die Ausbildung der PIC.

In dieser Doktorarbeit vergleiche ich die Transduktionsfrequenz von PPT-modifizierten HFV-Vektoren mit denen von lentiviralen Vektoren in nichtteilenden und teilenden Lungenkarzinomepithelzellen. Hierbei wurden Methoden wie Klonierung, Transfektion, und Transduktion (wie auch weitere Methoden) angewendet. Im Gegensatz zu lentiviralen Vektoren konnten HFV-Vektoren sich nicht teilende Zellen in meinen Versuchen nicht effizient transduzieren (Shityakov und Rethwiln, unveröffentlicht). Trotz der Befunde, dass HFV-Vektoren sichere und effiziente Alternativen zu lentiviralen Vektoren darstellen, bestehen immer noch große Einschränkungen, diese HFV-basierten, retroviralen Vektoren für Gentherapien bei ruhenden Zellen einzusetzen.

Viele Versuche wurden unternommen, um mögliche, vielversprechende Moleküle, welche als Wirkstoffe für eine HIV-Therapie eingesetzt werden könnten, zu finden. Diese Moleküle können aus chemischen Substanzbibliotheken bezogen werden oder am Computer *in silico* entworfen und dann synthetisiert werden. Digitalisierte Strukturen können als Referenzen benutzt werden, um besser herauszufinden, wie diese Moleküle Typen diverse Enzyme blokieren könnten. Strukturbasiertes Wirkstoffdesign hat das Potential, viele Jahre und Geld an Entwicklungskosten einzusparen.

Nachdem die Kristallstruktur der HIV-kodierten Proteasen aufgeklärt war, spielte das computergestützte Wirkstoffdesign eine zentrale Rolle bei der Entwicklung neuer Wirkstoffe gegen die Protease. Vielversprechende Vertreter dieser Wirkstoffklasse werden seit kurzem nun auch für die Behandlung von Patienten eingesetzt. Proteaseinhibitoren zeigen eine wirkungsvolle und langanhaltende Inhibition der HIV-1-Replikation; besonders dann, wenn sie in Kombinationstherapien eingesetzt werden. Aber diese Wirkstoffe werden immer weniger effektiv, je resistenter die HIV-Stämme durch Mutationen in den retroviralen Proteasen werden.

Im Rahmen meiner Arbeit mit computergestütztem Wirkstoffdesign nutzte ich Modellierungsmethoden wie den „lead expansion algorithm“ (Tripos®) um virtuelle Wirkstoffbibliotheken mit verschiedenen Affinitäten zur Proteasebindungsstelle zu erstellen. Zusätzlich wandte ich Verfahren der computergestützten, kombinatorischen Chemie an, um virtuelle Bibliotheken von Proteaseinhibitoren zu designen, und zu verbessern. Parallel dazu wurde eine *in silico* Selektion sowie eine Einteilung nach Pharmakophorähnlichkeiten für diese Kandidaten vorgenommen. Weiterführende computergestützte Analysen förderten einen einzigartigen Wirkstoff zu Tage, welcher neuartige Proteasebindungseigenschaften aufweist, und dessen Rolle für eine potentiell neuartige Klasse von Proteaseinhibitoren schon beschrieben wurde (Shityakov und Dandekar, 2009).

Eine Reihe von Modellen mit atomarer Auflösung wurden bereits entwickelt, um das Verhalten von Nanoröhren in Lipid-Doppelschichten aufzuklären.

Die Auswirkungen auf die molekular Dynamik einer einschichtigen Karbonnanoröhre, welche in das Zentrum einer Lipid-Doppelschicht eingefügt wurde, wurden intensiv studiert und analysiert. Die Normalabweichung und Fluktuationen wurden berechnet, um eine Aussage über die Stabilität der Lipid-Doppelschichten treffen zu können.

Die Ergebnisse weisen darauf hin, dass eine eingefügte Karbonnanoröhre die Freiheit für Konformationsänderungen bei nahegelegenen Lipiden einschränkt und dadurch einen Einfluss auf die Membranstabilität hat (Shityakov und Dandekar, 2011). Es kann aber außerdem sein, dass verschiedene Lipid-Doppelschichten Unterschiede in ihrer Fähigkeit, Brücken zwischen benachbarten Lipiden auszubilden, aufweisen.

Viren und Karbonnanoröhren werden damit in verschiedenen dynamischen Simulationen untersucht, um mehr über ihre Interaktionen mit Proteinen und Membranen zu erfahren.

Acknowledgments

I express my most gratitude to my advisors: Prof. Dr. Thomas Dandekar, head of the department of bioinformatics in Biocenter at the University of Würzburg and Prof. Dr. Axel Rethwilm, head of the Institute of Virology and Immunobiology at the University of Würzburg, for their valuable guidance, support, encouragement and patience during my doctoral studies at the University of Würzburg.

I am grateful to Dr. Karin Schleinkofer, Alexander Cecil and Santosh Nilla from the department of bioinformatics in Biocenter at the University of Würzburg; Dr. Tatiana Wiktorowicz from the Institute of Virology and Immunobiology at the University of Würzburg; Dr. Rob Russell from the department of structural bioinformatics, at the European molecular biology laboratory (EMBL) in Heidelberg.

I acknowledge the financial support from the Interdisziplinären Zentrum für Klinische Forschung (IZKF).

My heart feels thanks and gratitude to my parents for their understanding and encouragement. I wish to express my special gratitude and sincerest thank to the Almighty for giving me the strength and patience to overcome all the hardship during my postgraduate studies.

List of references

Achong, B.G., Mansell, W.A., Epstein, M.A., 1971. An unusual virus in cultures from a human nasopharyngeal carcinoma. J. Natl. Cancer Inst., 46:299-307.

An, D.G., Hyun, U., Shin, C.G., 2008. Characterization of nuclear localization signals of the prototype foamy virus integrase. J. Gen. Virol., 89(7):1680-1684.

Berg, B.A., 2004. Markov Chain Monte Carlo Simulations and Their Statistical Analysis (With Web-Based Fortran Code). World Scientific Publishing.

Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.H., Weissig, H., Shindyalov, I.H., Bourne, P.E., 2000. The protein data bank. Nucleic Acids Res., 28(1):235-242

Bieniasz, P.D., Weiss, R.A., McClure, M.O., 1995. Cell cycle dependence of foamy viruses. J. Virol., 69(11):7295-7299.

Bouyac-Bertoia, M., Dvorin, J.D., Fouchier, R.A., Jenkins, Y., Meyer, B.E., Wu, L.I., Emerman, M., Malim, M.H., 2001. HIV-1 infection requires a functional integrase NLS. Mol. Cell., 7(5):1025-1035.

Bushman, F.D., Fujiwara, T., Craigie, R., 1990. Retroviral DNA integration directed by HIV integration protein *in vitro*. Science., 249(4976):1555-1558.

Bowerman, B., Brown, P.O., Bishop, J.M., 1989. A nucleoprotein complex mediates the integration of retroviral DNA. Genes Dev., 3:469-478.

Brown, P.O., Bowerman, B., Varmus, H.E., 1989. Retroviral integration: structure of the initial covalent product and its precursor, and a role for the viral IN protein. Proc. Natl. Acad. Sci. USA., 86:2525-2529.

Bukrinsky, M., Haggerty, S., Dempsey, M., Sharova, N., Adzhubel, A., Spitz, L., Lewis, P., Goldfarb, D., Emerman, M., Stevenson, M., 1993. A nuclear localization signal within HIV-1 matrix protein that governs infection of non-dividing cells. Nat., 365:666-669.

Bieniasz, P. D., O. Erlwein, A. Aguzzi, A. Rethwilm, and M. O. McClure. 1997. Gene transfer using replication-defective human foamy virus vectors. Virology., 235:65-72.

Bianco, A., Hoebeke, J., Godefroy, S., Chaloin, O., Pantarotto, D., Briand, J.P., Muller, S., Prato, M., Partidos, C.D., 2005. Cationic carbon nanotubes bind to CpG oligodeoxynucleotides and enhance their immunostimulatory properties. J. Am. Chem. Soc., 127(1):58-59.

Bond, S.D., Benedict, J.L., Laird, B.B., 1999. The Nose-Poincare Method for Constant Temperature Molecular Dynamics. Comp. Phys., 151:114-134.

Balabin, I., Membrane Plugin, Version 1.1; 2008

Benson, N.C., Daggett, V., 2008. Dynameomics: large-scale assessment of native protein flexibility. Prot. Sci., 17(12):2038-2050.

Coman, R.M., Robbins, A.H., Fernandez, M.A., Gilliland, C.T., Sochet, A.A., Goodenow, M.M., McKenna, R., Dunn, B.M., 2008. The contribution of naturally occurring polymorphisms in altering the biochemical and structural characteristics of HIV-1 subtype C protease. Biochem., 47(2):731-743.

Charneau, P., Clavel, F.A., 1991. Single-stranded gap in human immunodeficiency virus unintegrated linear DNA defined by a central copy of the polypurine tract. J. Virol., 65:2415-2421.

Charneau, P., Alizon, M., Clavel, F.A., 1992. Second origin of DNA plus-strand synthesis is required for optimal human immunodeficiency virus replication. J. Virol., 66:2814-2820.

Charneau, P., Mirambeau, G., Roux, P., Paulous, S., Buc, H., Clavel, F., 1994. HIV-1 reverse transcription: a termination step at the center of the genome. J. Mol. Biol., 241:651-662.

Condra, J.H., Holder, D.J., Schleif, W.A., Blahy, O.M., Danovich, R.M., Gabryelski, L.J., Graham D.J., Laird, D., Quintero, J.C., Rhodes, A., Robbins, H.L., Roth, E., Shivaprakash, M.,Yang, T., Chodakewitz, J.A., Deutsch, P.J., Leavitt, R.Y., Massari, F.E., Mellors, J.W., Squires, K.E., Steigbigel, R.T., Teppler, H., Emini, E.A., 1996. Genetic correlates of *in vivo* viral resistance to Indinavir, a human immunodeficiency virus type 1 protease inhibitor., J. Virol., 70(12):8270-8276.

Condra, J.H., Schleif, W.A., Blahy, O.M., Gabryelski, L.J., Graham, D.J., Quintero, J.C., Rhodes, A., Robbins, HL., Roth, E., Shivaprakash, M., 1995. *In vivo* emergence of HIV-1 variants resistant to multiple protease inhibitors., Nat., 374(6522):569-571.

Poveda, E., de Mendoza, C., Martin-Carbonero, L., Corral, A., Briz, V., González-Lahoz, J., Soriano, V., 2007. Prevalence of darunavir resistance mutations in HIV-1-infected patients failing other protease inhibitors. J. Antimicrob. Chemother., 60(4):885-888.

Cane, P.A., de Ruiter, A., Rice, P., Wiselka, M., Fox, R., Pillay, D., 2001. Resistance-associated mutations in the human immunodeficiency virus type 1 subtype c protease gene from treated and untreated patients in the United Kingdom. J. Clin. Microbiol., 39(7):2652-2654.

Cramer, G.M., Ford, R.A., Hall, R.L., 1978. Estimation of toxic hazard - a decision tree approach. Food Cosmet. Toxicol., 16(3):255-276.

Connolly, M.L., 1983. Solvent-accessible surfaces of proteins and nucleic acids. Science, 221:709-713.

Connolly, M.L., 1983. Analytical molecular surface calculation. J. Appl. Crystallogr., 16:548-558.

Chook, Y.M., Blobel, G., 1999. Structure of the nuclear transport complex karyopherin-beta2-Ran x GppNHp. Nat., 299(6733):230-237.

Christ, F., Thys, W., De Rijck, J., Gijsbers, R., Albanese, A., Arosio, D., Emiliani, S., Rain, J.C., Benarous, R., Cereseto, A., Debyser, Z., 2008. Transportin-SR2 imports HIV into the nucleus. Curr. Biol., 18(16):1192-1202.

Creighton, T.E., 1993. Proteins: Structures and Molecular Properties. Freeman, San Francisco.

Cherukuri, P., Bachilo, S.M., Litovsky, S.H., Weisman, RB. Near-infrared fluorescence microscopy of single-walled carbon nanotubes in phagocytic cells. J. Am. Chem., 126(48):15638-15639.

Ceccarelli, M., Marchi, M., 1998. Molecular dynamics simulation of POPC at low hydration near the liquid crystal phase transition. Biochimie., 80(5-6):415-419.

Crowe, J.H., Crowe, L.M., Carpenter, J.F., Rudolph, A.S., Wistrom, C.A., Spargo, B.J., Anchordoguy, T.J., 1988. Interactions of sugars with membranes, Biochim. Biophys. Acta., 947:367-384.

Dorsey, B.D., Levin, R.B., McDaniel, S.L., Vacca, J.P., Guare, J.P., Darke, P.L., Zugay, J.A., Emini, E.A., Schleif, W.A., Quintero, J.C., 1994. The design of a potent and orally bioavailable HIV protease inhibitor. J. Med. Chem., 37(21):3443-3451.

Dubrovsky, L., Ulrich, P., Nuovo, G.J., Manogue, K.R., Cerami, A., Bukrinsky, M., 1995. Nuclear localization signal of HIV-1 as a novel target for therapeutic intervention. Mol. Med., 1(2):217-230.

Dresselhaus, M.S., Dresselhaus, G., Saito, R., 1995. Physics of carbon nanotubes, 883-891.

Gehlhaar, D.K., Verkhivker, G., Rejto, P.A., Fogel, D.B., Fogel, L.J., 1995. Docking Conformationally Flexible Small Molecules into a Protein Binding Site through Evolutionary Programming. Proceedings of the Fourth International Conference on Evolutionary Programming, 615-627.

Duhovny, D., Nussinov, R., Wolfson, H.J., 2002 Efficient unbound docking of rigid molecules. Proceedings of the Fourth International Workshop on Algorithms in Bioinformatics; Springer-Verlag GmbH, 185-200.

Douroumis, D., Fatouros, D.G., Bouropoulos N., Papagelis, K., Tasis, D., 2007. Colloidal stability of carbon nanotubes in an aqueous dispersion of phospholipids. Int. J. Nanomedicine., 2(4):761-766.

Dyda, F., Hickman, A.B., Jenkins, T.M., Engelman, A., Craigie, F., Davies, D.R., 1994. Crystal structure of the catalytic domain of HIV-1 integrase: similarity to other polynucleotidyl transferases. Sci., 266(5193):1981-1986.

Emiliani, S., Mousnier, A., Busschots, K., Maroun, M., Van Maele, B., Tempé, D., Vandekerckhove, L., Moisant, F., Ben-Slama, L., Witvrouw, M., Christ, F., Rain J.C., Dargemont, C., Debyser, Z., Benarous, R., 2005. Integrase mutants defective for interaction with LEDGF/p75 are impaired in chromosome tethering and HIV-1 replication. J. Biol. Chem., 280(27):25517-25523.

Engelman, A., Mizuuchi, K., Craigie, R., 1991. HIV-1 DNA integration: mechanism of viral DNA cleavage and DNA strand transfer. Cell., 67(6):1211-1221.

Feng, Y., Chen, M., Manley, J.L., 2008. Phosphorylation switches the general splicing repressor SRp38 to a sequence-specific activator. Nat. Struct. Mol. Biol., 15(10):1040-1048.

Gagner, J., Johnson, H., Watkins, E., Li, Q., Terrones, M., Majewski, J., 2006. Carbon nanotube supported single phospholipid bilayer. Langmuir., 22(26):10909-10911.

Garate, J.A., English, N.J., MacElroy, J.M., 2009. Static and alternating electric field and distance-dependent effects on carbon nanotube-assisted water self-diffusion across lipid membranes J. Chem. Phys., 131(11):114508.

Gallay, P., Hope, T., Chin, D., Trono, D., 1997. HIV-1 infection of nondividing cells through the recognition of intergase integration by the importin/kariopherin pathway. Proc. Natl. Acad. Sci. USA., 94(18):9825-9830.

Goodsell, D.S., Morris, G.M., Olson, A.J., 1996 Automated docking of flexible ligands: applications of AutoDock. J. Mol. Recognit., 9(1):1-5.

Goodsell, D.S., Olson, A.J., 1990. Automated docking of substrates to proteins by simulated annealing. Prot., 8(3):195-202.

Gunby, R.H., Sala, E., Tartari, C.J., Puttini, M., Gambacorti-Passerini, C., Mologni, L., 2007. Oncogenic fusion tyrosine kinases as molecular targets for anti-cancer therapy. Anti-Cancer Agents in Med. Chem., 7(6):594-611.

Hooks, J. J., Gibbs, C.J.Jr., 1975. The foamy viruses. Bacteriol. Rev., 39:169-185.

Brown, P., Nemo, G., Gajdusek D.C., 1978. Human foamy virus: further characterization, seroepidemiology, and relationship to chimpanzee foamy viruses. J. Infect. Dis., 137:421-427.

Herchenroder, O., Renne, R., Loncar, D., Cobb, E.K., Murthy, K.K., Schneider, J., Mergia, A., Luciw, P.A., 1994. Isolation, cloning, and sequencing of simian foamy viruses from chimpanzees (SFVcpz): high homology to human foamy virus (HFV). Virol., 201:187-199.

Heneine, W., Switzer, W.M., Sandstrom, P., Brown, J., Vedapuri, S., Schable, C.A., Khan, A.S., Lerche, N.W., Schweizer, M., Neumann-Haefelin, D., Chapman, L.E., Folks, T.M., 1998. Identification of a human population infected with simian foamy viruses. Nat. Med., 4:403-407.

Hill, C.L., Bieniasz, P.D., McClure, M.O., 1999. Properties of human foamy virus relevant to its development as a vector for gene therapy. J. Gen. Virol., 80:2003-2009.

Hooks, J.J., Gibbs, C.J. Jr., 1975. The foamy viruses. Bacteriol. Rev., 39:169-185.

Heinkelein, M., Dressler, M., Jarmy, G., Rammling, M., Imrich, H., Thurow, J., Lindemann, D., Rethwilm, A., 2002. Improved primate foamy virus vectors and packaging constructs. J. Virol., 76:3774-3783.

Henikoff, S., Henikoff, J.G., 1992. Amino acid substitution matrices from protein blocks. Proc. Natl. Acad. Sci. USA., 89(22):10915-10924.

Huber, H.E., Richardson, C.C., 1990. Processing of the primer for plus strand DNA synthesis by human immunodeficiency virus type-I reverse transcriptase. J. Biol. Chem., 265:10565-10573.

Humphrey, W., Dalke, A., Schulten, K., 1996. VMD: Visual molecular dynamics J. Molec. Graphics, 14:33-38.

Iijima, M., Suzuki, M., Tanabe, A., Nishimura, A., Yamada, M., 2006. Two motifs essential for nuclear import of the hnRNP A1 nucleocytoplasmic shuttling sequence M9 core. FEBS Lett., 580(5):1365-1370.

Imasaki, T., Shimizu, T., Hashimoto, H., Hidaka, Y., Kose, S., Imamoto, N., Yamada, M., Sato, M., 2007 Structural basis for substrate recognition and dissociation by human transportin 1. Mol. Cell, 28(1):57-67.

Imrich, H., Heinkelein, M., Herchenröder, O., Rethwilm, A., 2000. Primate foamy virus Pol proteins are imported into the nucleus. J. Gen. Virol., 81(12):2941-2947.

Izaurralde, E., Kutay, U., von Kobbe, C., Mattaj, I.W., Görlich, D., 1997. The asymmetric distribution of the constituents of the Ran system is essential for transport into and out of the nucleus. EMBO J., 16(21):6536-6547.

Ikeda, T., Nishitsuji, H., Zhou, X., 2004. Evaluation of the functional involvement of human immunodeficiency virus type 1 integrase in nuclear import of viral cDNA during acute infection. J. Virol., 78:11563-11573.

Jans, D.A., Hübner, S., 1996. Regulation of protein transport to the nucleus: central role of phosphorylation. Physiol. Rev., 76(3):651-685.

Kann, M., Sodeik, B., Vlachou, A., Gerlich, W.H., Helenius, A., 1999. Phosphorylation-dependent binding of hepatitis B virus core particles to the nuclear pore complex. J. Cell. Biol., 145(1):45-55.

Kam, N.W.S., Liu, Z., Dai, H., 2006. Functionalization of carbon nanotubes via cleavable disulfide bonds for efficient intracellular delivery of siRNA and potent gene silencing. J. Am. Chem. Soc., 127:12492-12493.

Kam N.W.S., Liu Z., Dai, H., 2006. Carbon Nanotubes as Intracellular Transporters for Proteins and DNA: An investigation of the uptake mechanism and pathway. Angew. Chem., Int. Ed., 45:577-581.

Kupiec, J.J., Tobaly, J., Canivet, M., Santillana, H.M., Flugel, R.M., Peries, J., Emanoil, R.R., 1988. Evidence for a gapped linear duplex DNA intermediate in the replicative cycle of human and simian spumaviruses. Nucleic Acids Res., 16:9557-9565.

Kupiec, J.J., Kay, A., Hayat, M., Ravier, R., Peries, J., Galibert, F., 1991. Sequence analysis of the simian foamy virus type I genome. Gene, 101:185-194.

Jaskólski, M., Tomasselli, A.G., Sawyer, T.K., Staples, D.G., Heinrikson, R.L., Schneider, J., Kent, S.B., Wlodawer, A., 1991. Structure at 2.5-A resolution of chemically synthesized human immunodeficiency virus type 1 protease complexed with a hydroxyethylene-based inhibitor. Biochem., 30(6):1600-1609.

Kutzner, C., Van der Spoel, D., Fechner, M.D., Lindahl, E., Schmitt, U.W., De Groot, B.L., Grubmüller, H., 2007. Speeding up parallel GROMACS on high-latency networks. J. Comp. Chem., 28: 2075-2084.

Koster, K.L., Webb, M.S., Bryant, G., Lynch, D.V., 1994. Interactions between soluble sugars and POPC (1-palmitoyl-2-oleoylphosphatidylcholine) during dehydration: vitrification of sugars alters the phase behaviour of the phospholipid. Biochim. Biophys. Acta., 1193(1):143-150.

Kohl, N.E., Emini, E.A., Schleif, W.A., Davis, L.J., Heimbach, J.C., Dixon, R.A., Scolnick, E.M., Sigal, I.S., 1988. Active human immunodeficiency virus protease is required for viral infectivity. Proc. Natl. Acad. Sci. USA., 85(13):4686-4690.

Kempf, D.J., Isaacson, J.D., King, M.S., Brun, S.C., Xu, Y., Real, K., Bernstein, B.M., Japour, A.J., Sun, E., Rode, R.A., 2001. Identification of genotypic changes in human immunodeficiency virus protease that correlate with reduced susceptibility to the protease inhibitor lopinavir among viral isolates from protease inhibitor-experienced patients. J. Virol., 75(16):7462-7469.

Lai, M.C., Kuo, H.W., Chang, W.C., Tarn, W.Y., 2003. A novel splicing regulator shares a nuclear import pathway with SR proteins. EMBL J., 22(6):1359-1369.

Lambert, C., Leonard, N., DeBolle, X., Depiereux, E., 2002. ESyPred3D: Prediction of proteins 3D structures. Bioinform., 18(9):1250-1256.

Linial, M., 2000. Why aren't foamy viruses pathogenic? Trends Microbiol., 8:284-289.

Luo, G., Sharmeen, L., Taylor, J., 1990. Specificities involved in the initiation of retroviral plus-strand DNA. J. Virol., 64:592-297.

Lavigne, M., Roux, P., Buc, H., Schaeffer, F., 1997. DNA curvature controls termination of plus strand DNA synthesis at the centre of HIV-1 genome. J. Mol. Biol., 266:507-524

Lu, R., Limon, A., Devroe, E., 2004. Class II integrase mutants with changes in putative nuclear localization signals are primarily blocked at a postnuclear entry step of human immunodeficiency virus type 1 replication. J. Virol., 78:12735-12746.

Lu, Q., Moore, J.M., Huang, G., Mount, A.S., Rao, A.M., Larcom, L.L., Ke, P.C., 2004. RNA polymer translocation with single-walled carbon nanotubes. Nano Lett., 4:2473-2477.

Liu, B., Li, X., Li, B., Xu, B., Zhao, Y., 2009. Carbon nanotube based artificial water channel protein: membrane perturbation and water transportation. Nano Lett., 9(4):1386-1394.

Levitt, M., 2001. The birth of computational structural biology, Nat. Struct. Biol., 8(5):392-393.

Martin, S.L., Garfinkel, D.J., 2003. Survival strategies for transposons and genomes. Genome Biol., 4(4):313.

Mattaj, I.W., Englmeier, L., 1998. Nucleocytoplasmic transport: the soluble phase. Annu. Rev. Biochem., 67:265-306.

Moll, A., Hildebrandt, A., Lenhof, H.P., Kohlbacher, O., 2005. BALLView: an object-oriented molecular visualization and modeling framework. J. Comput. Aided Mol. Des., 19(11):791-800.

Mikovits, J.A., Hoffman, P.M., Rethwilm, A., 1996. *In vitro* infection of primary and retrovirus-infected human leukocytes by human foamy virus. J. Virol., 70(5):2774-2780.

Meiering, C.D., Linial, M.L., 2003. The promyelocytic leukemia protein does not mediate foamy virus latency *in vitro*. J. Virol., 77(3):2207-2213.

Michalewicz, Z., 1992. Genetic Algorithms + Data Structures = Evolution Programs. Springer-Verlag.

Michalewicz, Z., Deb, K., Schmidt, M., Stidsen, T., 2000. Test - case generator for constrained parameter optimization techniques, IEEE Transactions on Evolutionary Computation.

Meiering, C. D., Linial, M.L., 2001. Historical perspective of foamy virus epidemiology and infection. Clin. Microbiol. Rev., 14:165-176.

Mergia, A., Leung, N.J., Blackwell, J., 1996. Cell tropism of the simian foamy virus type 1 (SFV-1). J. Med. Primatol., 25:2-7.

Mergia, A., Chari, S., Kolson, D.L., Goodenow, M.M., Ciccarone, T., 2001. The efficiency of simian foamy virus vector type-1 (SFV-1) in nondividing cells and in human PBLs. Virol., 280:243-252.

Meek, T.D., Dayton, B.D., Metcalf, B.W., Dreyer, G.B., Strickler, J.E., Gorniak, J.G., Rosenberg, M., Moore, M.L., Magaard, V.W., Debouck, C., 1989. Human immunodeficiency virus 1 protease expressed in Escherichia coli behaves as a dimeric aspartic protease. Proc. Natl. Acad. Sci. USA., 86(6):1841-1845.

Marrink, S.J., Berendsen, H.J.C., 1996 Permeation process of small molecules across lipid membranes studied by molecular dynamics simulations. J. Phys. Chem., 100:16729-16738, 1996.

Marrink S.J., Berger, O., Tieleman, P., Jahnig F., 1998. Adhesion forces in lipid membranes. Bioph. J., 74:931-943.

Nielsen, S.O., Ensing, B., Ortiz, V., Moore P.B., Klein, M.L., 2005. Lipid bilayer perturbations around a transmembrane nanotube: a coarse grain molecular dynamics study. Biophys. J., 88(6):3822-3828.

Nemo, G. J., Brown, P.W., Gibbs, C.J.Jr., Gajdusek, D.C., 1978. Antigenic relationship of human foamy virus to the simian foamy viruses. Infect. Immun., 20:69-72.

Needleman, S.B., Wunsch, C.D., 1970. A general method applicable to the search for similarities in the amino acid sequence of two proteins. J. Mol. Biol., 48(3):443-453.

Okada, M., Ye, K., 2009. Nuclear phosphoinositide signaling regulates messenger RNA export. RNA Biol., 6(1):12-16.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., Ferrin, T.E., 2004. UCSF Chimera - A Visualization System for Exploratory Research and Analysis. J. Comput. Chem., 25(13):1605-1612.

Pollard, V.W., Michael, W.M., Nakielny, S., Siomi, M.C., Wang, F., Dreyfuss, G., 1996. A novel receptor-mediated nuclear protein import pathway. Cell, 86(6):985-994.

Popov, S., Rexach, M., Zybarth, G., Reiling, N., Lee, M.A., L. Ratner, L., Lane, C.M., Moore, M.S., Blobel, G., Bukrinsky, M., 1998. Viral protein R regulates nuclear import of the HIV-1 pre-integration complex. EMBO J., 17:909-917.

Pirok, G., Maté, N., Varga, J., Szegezdi, J., Vargyas, M., Dórant, S., Csizmadia, F., 2006. Making "real" molecules in virtual space. J. Chem. Inf. Model., 46(2):563-568.

Petrek, M., Otyepka, M., Banás, P., Kosinová, P., Koca, J., Damborský, J., 2006. CAVER: a new tool to explore routes from protein clefts, pockets and cavities. BMC Bioinf., 7:316.

Pullen, A.K., Rattray, A.J., Champoux, J.J., 1993. The sequence features important for plus strand priming by human immunodeficiency virus type 1 reverse transcriptase. J. Biol. Chem., 268:6221-6227.

Perryman, A.L., Lin, J.H., McCammon, J.A., 2004. HIV-1 protease molecular dynamics of a wild-type and of the V82F/I84V mutant: possible contributions to drug resistance and a potential new target site for drugs. Prot. Sci., 13(4):1108-1123.

Peter, C., Hummer, G., 2005. Ion transport through membrane-spanning nanopores studied by molecular dynamics simulations and continuum electrostatics calculations. Biophys. J., 89:2222-2234.

Pan, Y., MacKerell, A.D.Jr., 2003.Altered structural fluctuations in duplex RNA versus DNA: a conformational switch involving base pair opening. Nucleic Acids Res., 31(24):7131-7140.

Priyakumar, U.D., Ramakrishna, S., Nagarjuna, K.R., Reddy, S.K., 2010. Structural and energetic determinants of thermal stability and hierarchical unfolding pathways of hyperthermophilic proteins, Sac7d and Sso7d. J. Phys. Chem. B., 114(4):1707-1718.

Quinn, P.J., 1989. Principles of membrane stability and phase behaviour under extreme conditions. J. Bioenerg. Biomembr., 21:3-19.

Qiao, R., Ke, P.C., Lipid-carbon nanotube self-assembly in aqueous solution. J. Am. Chem. Soc., 128(42):13656-13657.

Rebane, A., Aab, A., Steitz, J.A., 2004. Transportins 1 and 2 are redundant nuclear import factors for hnRNP A1 and HuR. RNA, 10(4):590-599.

Rio, D.C., 1991. Regulation of Drosophila P element transposition. Trends Genet., 7(9):282-287.

Richardson, J.S., Schneider, B., Murray, L.W., Kapral, G.J., Immormino, R.M., Headd, J.J., Richardson, D.C., Ham, D., Hershkovits, E., Williams, L.D., Keating, K.S., Pyle, A.M., Micallef, D., Westbrook, J., Berman, H.M., 2008. RNA backbone: consensus all-angle conformers and modular string nomenclature (RNA Ontology Consortium contribution). RNA, 14(3):465-481.

Roe, T., Reynolds, T.C., Yu, G., Brown, P.O., 1993. Integration of murine leukemia virus DNA depends on mitosis. EMBO J., 12:2099-2108.

Russell, D. W., Miller, A.D., 1996. Foamy virus vectors. J. Virol. 70:217-222.

Richard, C., Balavoine, F., Schultz, P., Ebbesen, T.W., Mioskowski, C., 2003. Supramolecular self-assembly of lipid derivatives on carbon nanotubes. Sci., 300(5620):775-778.

Saïb, A., Puvion-Dutilleul, F., Schmid, M., Périès, J., de Thé, H., 1997. Nuclear targeting of incoming human foamy virus Gag proteins involves a centriolar step. J. Virol., 71(2):1155-1161.

Sturgeon, J.B., Laird, B.B., 2002. Symplectic algorithm for constant pressure molecular dynamics using a Nos-Poincar thermostat. University of Kansas technical paper.

Sun, G., 1993. Symplectic partitioned Runge-Kutta methods, J. Comput. Math., 11:365-372.

Stern, J., Feisleben H.J., Janku S., Ring K., 1989. Black lipid membranes of tetraether lipids from *Thermoplasma acidophilum*. Biochim. Biophys Acta., 1128:227-236.

Sens, P., Isambert, H., 2002. Undulation instability of lipid membranes under an electric field. Phys. Rev. Lett., 88(12):128102.

Schneidman-Duhovny D., Inbar Y., Nussinov R., Wolfson H.J., 2005. PatchDock and SymmDock: servers for rigid and symmetric docking. Nucleic Acids Res., 33:363-367.

Siomi, H., Dreyfuss, G., 1995. A nuclear localization domain in the hnRNP A1 protein. J. Cell. Biol., 129(3):551-560.

Schweizer, M., Neumann-Haefelin, D., 1995. Phylogenetic analysis of primate foamy viruses by comparison of pol sequences. J. Virol., 207:577-582.

Switzer, W.M., Bhullar, V., Shanmugam, V., 2004. Frequent simian foamy virus infection in persons occupationally exposed to nonhuman primates. J. Virol., 78(6):2780-2789.

Tarek, M., 2005.Membrane electroporation: a molecular dynamics simulation. Biophys. J., 88(6):4045-4053.

Thauvin, C., Rickling, S., Schultz, P., Célia, H., Meunier, S., Mioskowski, C., 2008. Carbon nanotubes as templates for polymerized lipid assemblies. Nat. Nanotechnol., 3(12):743-748.

Trobridge, G.D., Josephson, N., Vassilopoulos, G., Mac, J., Russell, D.W., 2002. Improved foamy virus vectors with minimal viral sequences. Mol. Ther., 6:321-328.

Trobridge, G. D., Russell, D.W., 1998. Helper-free foamy virus vectors. Hum. Gene Ther., 9:2517-2525.

Thomsen, R., Christensen, M.H., 2006. MolDock: a new technique for high-accuracy molecular docking. J. Med. Chem., 49(11):3315-3321.

Tanrikulu, Y., Nietert, M., Scheffer, U., Proschak, E., Grabowski, K., Schneider, P., Weidlich, M., Karas, M., Göbel, M., Schneider, G., 2007. Scaffold hopping by "fuzzy" pharmacophores and its application to RNA targets. Chembiochem., 8(16):1932-1936.

Temin, H.M., Baltimore, D., 1972. RNA-directed DNA synthesis and RNA tumor viruses. Adv. Virus Res., 17:129-186.Tomasselli, A.G., Hui, J.O., Sawyer, T.K., Staples, D.J., Bannow, C., Reardon, I.M., 1990. Scecificity and inhibition of proteases from human immunodeficiency virus 1 and 2, J. Biol. Chem., 265:14675-14683.

Tsvetkov, T.D., Tsonev, L.I., Tsvetkova, N.M., Koynova, R.D. Tenchov, B.G. 1989. Effect of trehalose on the phase properties of hydrated and lyophilized dipalmitoylphosphatidylcholine multilayers. Cryobiol., 26:162-169.

Tokunaga, K., Ikuta, K., Adachi, A., Matsuda, M., Kurata, T., Kojima, A., 1999. The cellular kinase binding motifs (PxxP and RR) in human immunodeficiency virus type 1 Nef protein are dispensable for producer-cell-dependent enhancement of viral entry. J. Virol., 257(2):285-289.

Vacca, J.P., Dorsey, B.D., Schleif, W.A., Levin, R.B., McDaniel, S.L., Darke, P.L., Zugay, J., Quintero, J.C., Blahy, O.M., Roth, E., 1995. An orally bioavailable human immunodeficiency virus type 1 protease inhibitor. Proc. Natl. Acad. Sci. USA., 91(9):4096-40100.

Vakifahmetoglu, H., Zhivotovsky, B., 2009. Do checkpoint kinases contribute to HIV-1-mediated apoptosis? Cell Cycle, 8(3):335-336.

Vassilopoulos, G., Trobridge, G., Josephson, N.C., Russell, D.W., 2001. Gene transfer into murine hematopoietic stem cells with helper-free foamy virus vectors. Blood., 98:604–609.

Van der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A.E., Berendsen, H.J.C., 2005. Gromacs: Fast, Flexible and Free. J. Comp. Chem., 26:1701-1718.

Von Schwedler, U., Kornbluth, R.S., Trono, D., 1994. The nuclear localization signal of the matrix protein of human immunodeficiency virus type 1 allows the establishment of infection in macrophages and quiescent T lymphocytes. Proc. Natl. Acad. Sci. USA., 91:6992-6996.

Vakarelski, I.U., Brown, S.C., Higashitani, K., Moudgil, B.M., 2007. Penetration of living cell membranes with fortified carbon nanotube tips. Langmuir., 23(22):10893-10896.

Willard, L., Ranjan, A., Zhang, H., Monzavi, H., Boyko, R.F., Sykes, B.D., Wishart, D.S., 2003. VADAR: a web server for quantitative evaluation of protein structure quality. Nucleic Acids Res., 31(13):3316-3319.

Willard, L., Ranjan, A., Zhang, H., Monzavi, H., Boyko, R.F., Sykes, B.D., Wishart, D.S., 2003. VADAR: a web server for quantitative evaluation of protein structure quality. Nucleic Acids Res., 31(13):3316-3319.

Wolber, G., Langer, T., 2005. LigandScout: 3-D pharmacophores derived from protein-bound ligands and their use as virtual screening filters. J. Chem. Inf. Model., 45(1):160-169.

Wlodawer, A., Miller, M., Jaskólski, M., Sathyanarayana, B.K., Baldwin, E., Weber, I.T., Selk, L.M., Clawson, L., Schneider, J., Kent, S.B., 1989. Conserved folding in retroviral proteases: crystal structure of a synthetic HIV-1 protease. Sci., 245(4918):616-621.

Winslow, D.L., Otto, M.J., 1995. HIV protease inhibitors. AIDS, 9:183-192.

Wallace, E.J., Sansom, M.S., 2009. Carbon nanotube self-assembly with lipids and detergent: a molecular dynamics study. Nanotechnol., 20(4):045101.

Wang, H., Michielssens, S., Moors, S.L.C., A. Ceulemans, A., 2009. Molecular dynamics study of dipalmitoylphosphatidylcholine lipid layer self-assembly onto a single-walled carbon nanotube. Nano Res., 2:945-954.

Wang, H., 2009. Dispersing carbon nanotubes using surfactants. Current Opinion in Colloid and Interface Sci., 14:364-371.

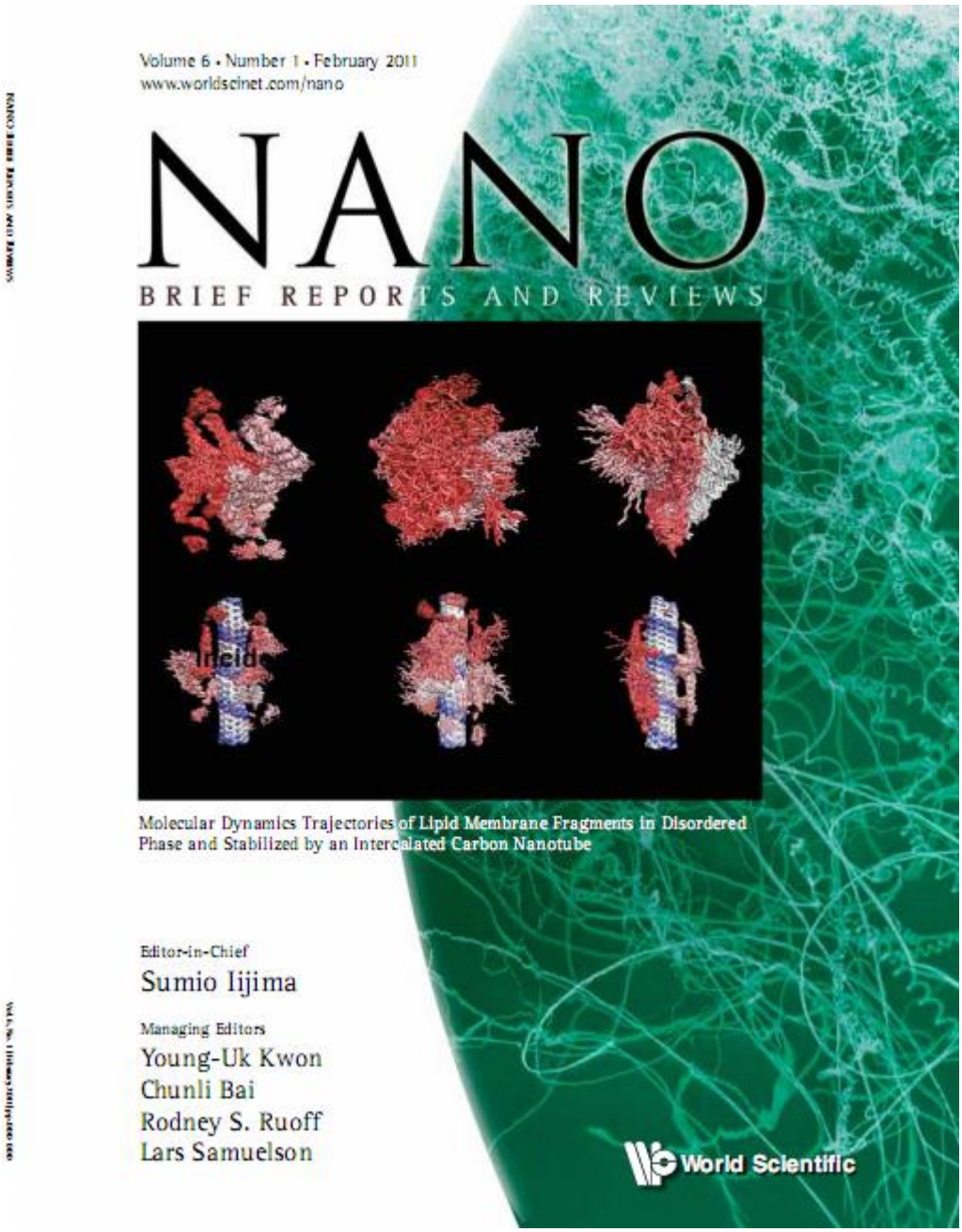
Wu, Y., Hudson, J.S., Lu, Q., Moore, J.M., Mount, A.S., Rao, A.M., Alexov, E., Ke, P.C., 2006. Coating single-walled carbon nanotubes with phospholipids. J. Phys. Chem. B., 110(6):2475-2478.

York, D.M., Darden, T.A., Pedersen, L.G., Anderson, M.W., 1993. Molecular dynamics simulation of HIV-1 protease in a crystalline environment and in solution. Biochem., 32(6):1443-53.

Yang, J.M., Chen, C.C., 2004. GEMDOCK: a generic evolutionary method for molecular docking., Prot., 55(2):288-304.

Zhou, X., Moran-Mirabal, J.M., Craighead, H.G., McEuen, P.L., 2007. Supported lipid bi-layer/carbon nanotube hybrids. Nat. Nanotechnol., 2(3):185-190.

Publications related to this work



1. **Shityakov, S*.**, Dandekar, T., 2011.

Molecular dynamics simulation of POPC and POPE lipid membrane bilayers enforced by an intercalated single-wall carbon nanotube.

NANO, Vol 6, No 1, (my simulations featured as front paper with the cover image)

2**. Shityakov, S***., Dandekar, T*., Rethwilm, A., 2010.

Structural and docking analysis of HIV-1 integrase and Transportin-SR2 interaction: Is this a more general and specific route for retroviral nuclear import and its regulation?
Online Journal of Bioinformatics, Volume 11 (1): 19-33,

3. **Shityakov S***., Dandekar , T., 2009.
Lead expansion and virtual screening of Indinavir derivate HIV-1 protease inhibitors using pharmacophoric - shape similarity scoring function.
Bioinformation, 2010, Vol. 4, No. 7, pp. 295-299

*-Corresponding Author

Conference contributions and participations

1. Poster: High temperature molecular dynamics simulation of POPC and POPE lipid membrane bilayers enforced by an intercalated single-wall carbon nanotube
Attendance Certificate: CHIASMA 5th International, Würzburg, Germany, 13-14 October, 2010

2. Presentation: Three-dimensional quantitative structure-activity relationship analysis and ADME predictions of guanylhydrazone coactivator binding inhibitors of estrogen receptors.
DGKL (Deutsche Vereinte Gesellschaft für Klinische Chemie und Labortoriummedizin) Wissenschaftlicher Kongress, Mannheim, Germany, 29 - 02 September/October, 2010

3. Presentation: High temperature molecular dynamics simulation of POPC and POPE lipid membrane bilayers enforced by an intercalated single-wall carbon nanotube
Conference on computational physics 2010 (CCP 2010), Trondheim, Norway 23-26 June 2010

4. Presentation: Analysis of the HIV-1 central polypurine tract in a foamy viral vector background
Attendance Certificate: 8th Foamy Virus Meeting, Argos, Greece 06 May 2010

5. *Attendance Certificate: The SFB 479 International Symposium, Würzburg, Germany, 16-18 July, 2009*

6. *Attendance Certificate: IZKF (Interdisziplinäres Zentrum für Klinische Forschung an der Julius-Maximilians-Universität) Seminar, Würzburg, Kloster Banz, Germany, 30-31 May, 2008*

Training courses, workshops and lectures

1. *Attendance Certificate: 'Clinical Medicine' lecture series, April 20- July 6, 2010.*

Topics included:

1) Gene therapy in neurodegenerative diseases
2) Osteoarthritis and articular cartilage repair
3) Multiple myeloma
4) Neuroimmunological diseases
5) Genetics of aging
6) Novel vaccination strategies against infectious diseases
7) Pulmonary hypertension
8) Vascular diseases/Atherosclerosis
9) Immune therapeutic approaches for neurodegenerative diseases
10) Functional brain imaging in psychiatry
11) Emotion, consciousness and the self

2. *Attendance Certificate: Workshop on 'Microcapillary bench-top Guava Flowcytometer', May 25, 2010.*

3. *Attendance Certificate: Training courses on 'Control of Infectious Tropical Disease', November 23-24, 2009.*

Topics of the training

Malaria – the basics:

1) Treatment and control of Malaria
2) Diagnostic methods under minimal conditions

3) Treatment and control of Malaria

Human African Trypanosomiasis (HAT)

1) Diagnostic methods of HAT

2) Treatment and control of HAT

Introduction of Leishmaniasis

1) Diagnostic methods of Leishmaniasis

2) Therapy and control of Leishmaniasis

Isolation unit for highly contagious diseases

Tuberculosis in countries with limited resources

1) Diagnostic methods of TB

2) Treatment and control of TB in resource poor settings

Erklärung

Hiermit erkläre ich ehrenwörtlich, daß ich die vorliende Dissertation selbsständig angefertigt und keine anderen als die angegeben Quellen und Hilfsmittel verwendet habe.

Die Dissertation wurde bische in gleicher noch änlicher Form in einem anderen Prüfungsverfahren vorgelegt.

Außer dem Diplom in Medizin der Russischen Medizinischen Staatsuniversität und den Dr. med. von der Bayrischen Julius-Maximilians Universität Würzburg habe ich bisher keine weiteren akademischen Graden erworben oder versucht zu erwerben.

Würzburg, 2011
Sergey Shityakov