Biocomputing approaches to protein-protein interactions in molecular medicine.

Von der Fakultät für Mathematik, Informatik und Naturwissenschaften der RWTH Aachen University zur Erlangung des akademischen Grades einer Doktorin der Naturwissenschaften genehmigte Dissertation

vorgelegt von

Valeria Losasso, PhD

aus Teramo, Italy

Berichter: Universitätprofessor Dr. Paolo Carloni
Universitätprofessor Dr. Marc Spehr

Tag der mündlichen Prüfung: 23.11.2012

Diese Dissertation ist auf den Internetseiten der Hochschulbibliothek online verfügbar.

Biocomputing approaches to protein-protein interactions in molecular medicine.



Valeria Losasso

Forschungszentrum Juelich

RWTH Aachen

A thesis submitted for the degree of

Doctor Rerum Naturalium

2012

1. Reviewer: Prof. Paolo Carloni

2. Reviewer: Prof. Marc Spehr

Abstract

Proteins rarely act alone. They often undertake biological functions as components of larger molecular machines characterized by intricate physicochemical dynamic connections. Current research efforts increasingly point to genome-wide screenings for protein complexes. These data have to be integrated with structural information in order to gain insights into functional aspects. In this context, computational studies play a crucial role, being able to complement experimental data. In this thesis, I have used computational methods to unravel key aspects of several pharmacologically relevant protein complexes, using increasingly sophisticated computational tools.

By using molecular modeling and classical molecular dynamics, I have predicted the atomic details of the protein complex formed by Granzyme B and its *in vivo* inhibitor Serpin B9. Granzyme B is a component of the immune system that could be exploited for the design of recombinant constructs effective in cancer therapy. The work has been carried out in collaboration with Prof. Barth at Fraunhofer Institute, RWTH - University of Aachen. Next, I have applied protein-protein docking procedures to explore the possible binding poses between PARP10 and GSK3β, two interacting proteins involved in cancer pathways. This work is in collaboration with Prof. Lüscher at the Clinics of RWTH Aachen University. Finally, in close collaboration with Dr. Adriana Pietropaolo (U. Catanzaro, Italy), I have investigated the conformational flexibility of the complex between the neurotrophin NGF, essential for neuron function and survival, and its receptor TrkA. This complex is relevant for Alzheimer's therapy. The results show that the structure of the complex in solution may differ from that in the solid state (so far the only one available). The previously unrecognized

structural information could be used to design of new NGF mimics and signaling modulators in Alzheimer's therapy.

To my family

Acknowledgements

I wish to thank my supervisor Prof. Paolo Carloni not only for the fruitful scientific discussions, but also for having put his trust in me even in the most difficult stages of my study. A very special thanks goes out to all of our collaborators, namely Prof. Stefan Barth, Sonja Schiffer, Prof. Bernhard Lüscher, Karla Fejis and Dr. Adriana Pietropaolo, whose contribution has been crucial for my work. I would also like to acknowledge the entire GRS group for their invaluable support during these years and all the administration staff for their daily work. Moreover, I really thank my family for the encouragement they provided me not only during this PhD but also during my entire life; all of my friends and Gianfranco, whose support was indispensable during the last year.

Contents

List of Figures **vii**

List of Tables **ix**

1 Introduction **1**

2 Molecular medicine topics investigated in this thesis. **7**

2.1 Cancer 7

2.1.1 Immunotherapy: immunotoxins 8

2.1.1.1 Granzyme B. 9

2.1.1.2 Serpin B9. 12

2.1.2 Targeted therapy: PARP inhibitors. 14

2.1.2.1 PARPs structure and function. 14

2.1.2.2 PARP10. 15

2.2 Neurotrophins in neurodegenerative disorders. 19

2.2.1 The Nerve Growth factor 22

2.2.2 The NGF/TrkA structure. 23

3 Methods **27**

3.1 Structural prediction from bioinformatics 27

3.1.1 Predictions for a single protein 27

3.1.1.1 Sequence alignment 27

3.1.2 Homology modeling 33

3.1.3 Protein-protein complex structural predictions. 38

3.1.3.1 Introduction. 38

3.1.3.2 Multistep procedure in docking. 39

3.1.3.3 Model evaluation. 41
3.2 Physics-based methods: Molecular dynamics and free energy calculations. 43
3.2.1 Introduction 43
3.2.2 Integration of the equations of motions. 45
3.2.3 Force field. 46
3.2.4 Periodic boundary conditions. 48
3.2.5 Neighbors list. 49
3.2.6 Long range interactions. 50
3.2.6.1 Ewald sum. 50
3.2.6.2 Particle Mesh Ewald. 51
3.2.7 Temperature and pressure control. 51
3.2.7.1 Berendsen thermostat. 52
3.2.7.2 Nose-Hoover temperature coupling. 53
3.2.7.3 Langevin thermostat. 53
3.2.7.4 Berendsen barostat. 53
3.2.7.5 Nose-Hoover / Langevin piston. 54
3.2.8 Free energy calculations. 55

4 Design of Human Granzyme B Variants resistant to Serpin B9. 57
4.1 Introduction. 57
4.2 Methods. 59
4.3 Results and discussion 63
4.4 Concluding remarks. 70

5 Investigation of PARP10-GSK3β interactions at the molecular level 71
5.1 Introduction 71
5.2 Computational methods. 75
5.2.1 Available structural information for the isolated moieties. 75
5.2.2 Docking NAD^+ onto PARP10. 76
5.2.3 Docking PARP10/NAD^+ onto GSK3β. 77
5.2.4 Experimental methods (Carried out by M.Sc. Karla Fejis in the lab of Prof. Lüscher at the Medical School of RWTH Aachen University). 77
5.3 Results. 78

5.3.1 Structural prediction PARP10-NAD$^+$. 78
5.3.2 Docking PARP10-GSK3β. 79
5.3.3 Mutations. 81
5.4 Discussion. 81

6 Molecular simulation of the NGF-TrkA complex. 83
6.1 Introduction 83
6.2 Methods. 85
6.2.1 Equilibration. 85
6.2.2 Free energy calculations. 85
6.3 Results 88
6.3.1 Structural features. 88
6.3.2 Validation with experimental data. 90
6.4 Discussion and conclusions. 90

7 Conclusions 95

A Design of Human Granzyme B Variants resistant to Serpin B9. 99

B NGF binding to p75 and downstream signaling pathways. 105

References 107

List of Figures

1.1 Granzyme B-Serpin B9 complex 3
1.2 GSK3β structure 4
1.3 NGF/TrkA conformers. 5

2.1 An example of immunotoxin. 8
2.2 Mechanism of action of GB. 10
2.3 GB structure. 11
2.4 Typical serpin fold. 13
2.5 Poly-ADP-ribose formation by PARPs. 15
2.6 Comparison between PARP1 and PARP10 structural organization. . . 16
2.7 Structure-based sequence alignment of PARP1, PARP3, PARP10, and PARP12 catalytic domains. 16
2.8 GSK3β structure. 18
2.9 Neurotrophin receptors. 20
2.10 Structure of Nerve Growth Factor Complexed with the Extracellular Domain of TrkA 24

3.1 The dot matrix technique for sequence alignment. 30
3.2 An example of dynamic programming for two protein sequences. 31
3.3 Relation between sequence and structure homology. 34
3.4 Steps in comparative protein structure modeling. 35
3.5 Schematic representation of periodic boundary conditions in 2D. 49

4.1 Sequence alignments. 60
4.2 Structural properties. 64
4.3 Protein-protein interfaces. 66

4.4 In vitro assays. 70

5.1 Superposition of PARP1 and PARP10 catalytic sites. 72
5.2 Identified PARP10 substrates. 73
5.3 PARP10 scorpion motif 76
5.4 PARP10-NAD$^+$ complex. 79
5.5 Docking poses 80

6.1 NGF structure complexed with TrkA extracellular domains (two side views). 84
6.2 Collective variables (θ_1, θ_2, θ_3) for the pincer movement of NGF/TrkA complex (40). 87
6.3 Free energy landscape of NGF/TrkA complex 89
6.4 CD spectra 90
6.5 NGF/TrkA minima A,B,C,D 92

A.1 Template for homology modeling. 101
A.2 Structure of wt hGB·hSB9 102
A.3 Structures of wt and mutated hGB·hSB9 complexes. 103

List of Tables

4.1 Structural Properties of wt and mutated complexes. 64
4.2 Contacts between hGB and hSB9. 65

5.1 Validated PARP10 substrates. 74

A.1 Alanine Scanning. 100
A.2 Details of the simulations. 100
A.3 Trypsin-Serpin B1 interactions. 102

LIST OF TABLES

1

Introduction

Protein-protein interactions maintain and regulate most processes in the cell, from metabolism to transcription, from growth, to differentiation and programmed cell death (1, 2, 3). System biology approaches[1] allow to predict a variety of properties of the resulting interactomes. Indeed, the systematic protein-protein interaction mapping, when coupled with experimental validation, provides insights into the properties of complexes integrated into their biological networks, as well as into the disease mechanisms at a system level (10). A remarkable result in this regard just appeared while writing this thesis. A whole-cell computational model of the human pathogen *M. genitalium* (11) turned out to account for all of the annotated gene functions and the related experimental data. This very recent achievement clearly shows that computations are greatly advancing our understanding of complex cellular behaviors (11).

In spite of its successes, however, the applicability and predictive power of systems biology is often limited by the lack of structural information (12). Indeed, because of the limitations of the experimental structural techniques (13)[2], a significant fraction of protein/protein complexes is expected to be extremely difficult to study and/or highly time consuming, using classical structural methods (16). In this context, using efficient

[1] These approaches are based on data coming from several experimental sources, including yeast-two hybrid systems (4), mass spectrometry (5), colocalisation (6), analysis of gene fusion (7), phylogenetic profiling (8) and co-evolution (9).

[2] Structural information at low-to-intermediate resolution by mass spectrometry, cryo-electron microscopy, and small-angle X-ray scattering (14) is usally complemented by high resolution NMR spectroscopy and X-ray crystallography methods (15).

computational approaches for protein complex structure predictions is crucial (17). In particular, when the structural determinants of two interacting proteins are known, computational docking represents a valuable strategy to predict the structure of their complex (18, 19, 20, 21, 22). In docking procedures, it is of enormous help to include experimental data (obtained for example from molecular biology (23) and NMR studies (24)) to guide the calculations. At times, it is also very useful to use template-based prediction algorithms, recently developed on the basis of known structural complexes, which combine structural similarity and evolutionary conservation in protein interfaces to infer structural details of uncharacterized complexes (25). Conversely, when the structure of the complex is known, one can use computational methods such as the identification of surface residues by combining structures with sequence alignments and phylogenetic trees (26, 27) or the comparative analysis of 'hot spots' identified by alanine scanning (28) to identify the determinants of protein interactions. Advanced molecular-dynamics based methods such as free energy calculations (29) have also made significant progresses in the past few decades (30), allowing one to describe the energetics of protein/protein complexes.

In this work, we have applied state-of-the-art computational approaches to challenging systems of pharmaceutical relevance, from innovative therapeutic strategies against cancer to protein/protein complexes playing a role for Alzheimer's disease.

In the context of anti-cancer therapy, I first investigated the interaction between human Granzyme B, a protein involved in immune-mediated apoptosis and thus relevant in the control of cancer development, and its *in vivo* inhibitor Serpin B9 (31). By predicting first the structure of Serpin B9 and then that of the complex (Fig. 1.1), I identified the structural determinants of this interaction to rationally design Granzyme B mutants. These featured a significantly decreased binding to Serpin B9 and an enzymatic activity comparable to that of the wild-type isoform. These mutants have been expressed by our collaborators, in the group of Prof. S. Barth at Fraunhofer Institute for Molecular Biology and Applied Ecology at RWTH - University of Aachen. They may be exploited for the design of recombinant constructs of therapeutical relevance able to specifically target tumor cells.

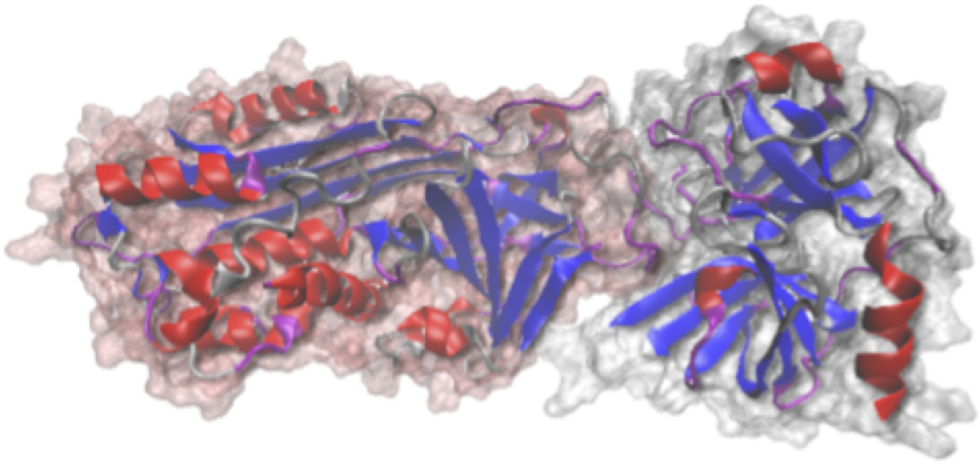

Figure 1.1: Granzyme B-Serpin B9 complex. Predicted complex between Granzyme B (white surface) and Serpin B9 (pink surface).

We next turned our attention to the interaction between the PARP10 and GSK3β proteins. Both are involved in cancer pathways, the first one by adding one ADP-ribose moiety on specific targets and thus promoting DNA repair (130), and the second one by modulating cell death through the interaction with key transcription factors such as c-MYC (33). Unlike most of the other members of its protein family, PARP10 lacks a glutamate of the characteristic catalytic triad "HYE" in the active site. Therefore, PARP10 likely acts by exploiting an exposed acidic residue of its interaction partner to carry out the enzymatic reaction (34). The involvement of lysines and/or arginines was also proposed (35, 36). We applied a protein-protein docking procedure to screen all of the possible GSK3β acidic and/or basic surface residues which may interact with PARP10 (Fig. 1.2). This study has been carried out in collaboration with the experimental group of Prof. B. Lüscher at the Medical School of the RWTH Aachen University. The results, although still preliminary, represent a step forward in the understanding of PARP10 mode of action.

Our final study is of relevance for Alzhemer's therapy. Here I provided a quantitative description of the energetics of a protein/protein complex in aqueous solution. This part has been carried out in close collaboration with Dr. A. Pietropaolo (University of Catanzaro, Italy). I focused on the nerve growth factor (NGF), a secreted protein which is crucial for neurons growth and survival, and its receptor TrkA (37). The

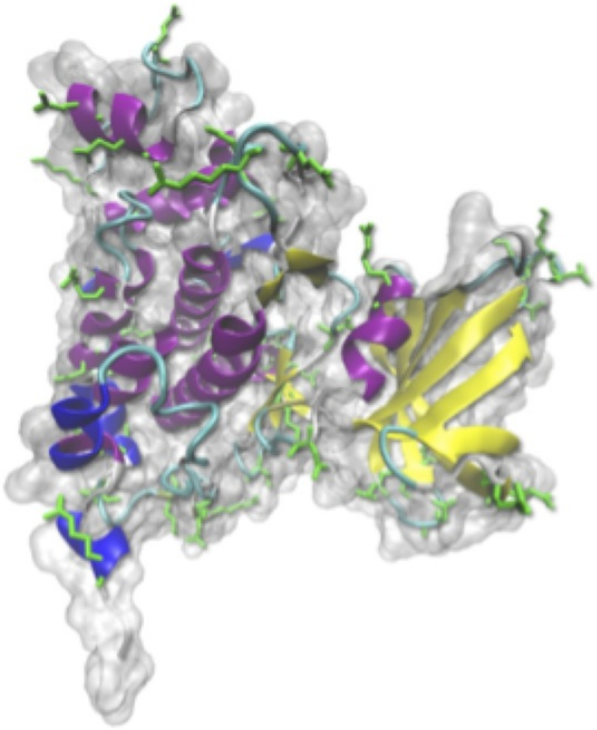

Figure 1.2: GSK3β structure. The exposed residues considered as potential PARP10 interactors are represented in green sticks.

NGF/TrkA downstream signaling has been involved in Alzheimer's disease pathogenesis (38). Indeed, the so-called trophic action exerted by NGF (and possibly also by other neurotrophins) in the target neurons has been shown to be actually the result of an anti-amyloidogenic activity, in that the NGF signaling interruption leads to an increased production of amyloid beta peptide (Abeta) associated to TrkA and to its intracellular aggregation (39). The structure of NGF in complex with TrkA has been solved by X-ray crystallography (40). Our calculations show that conformers other than that found in the X-ray structure are present in aqueous solution (Fig. 1.3).

These structures provide a molecular basis for molecular biology data and could be used for the rational design of TrkA signaling modulators with an improved accuracy.

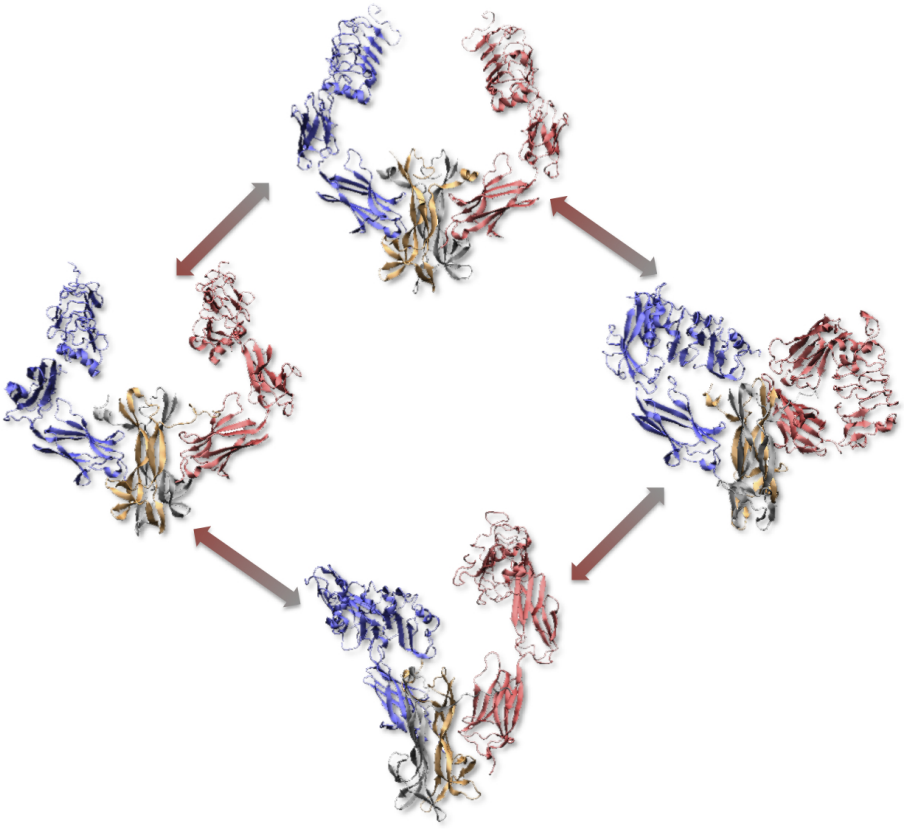

Figure 1.3: NGF/TrkA conformers. NGF/TrkA conformers present in aqueous solution according to our calculations. The two TrkA chains are in blue and red cartoons, the two NGF chains in grey and orange cartoons.

2

Molecular medicine topics investigated in this thesis.

2.1 Cancer

Cancer is a leading cause of death worldwide and accounted for 7.6 million deaths (13% of all deaths) in 2008 (41). It is a multi-factorial disease. Enviromental pollutants, infections and incorrect lifestyles may directly damage genes or combine with existing genetic faults within cells to cause the disease. A small percentage of cancers, approximately 5-10%, are entirely hereditary (42, 43). Its essential hallmarks are uncontrolled cell survival, growth, angiogenesis and metastasis (44).

Classical therapeutic strategies are represented by surgery, radiotherapy and chemotherapy. Novel strategies include: i) hormonal therapy, i.e. the use of drugs to block the effects of hormones used by the cancer to grow (45); ii) angiogenesis inhibition, achieved with the use of drugs inhibiting the growth of new blood vessels that could feed tumors (46); iii) immunotherapy, based on the activation of patients immune system at a multistep level (promoting the production of protective T-cell responses and overcoming immunosuppression in the tumour bed) (47); iv) targeted therapy, which uses drugs specifically directed against well-defined targets that are critical for tumor survival and not compromising for normal organs and tissues (48).
The last two strategies represent the background of the biological systems studied in this thesis.

2.1.1 Immunotherapy: immunotoxins

Immunotoxins (ITs) represent a very promising class of antitumor agents, being composed of a targeting moiety (a specific ligand or monoclonal antibody) linked to a cytotoxic moiety (49, 50). ITs bind to a surface antigen on a cancer cell, enter the cell by endocytosis, and kill it (Fig. 2.1) (49).

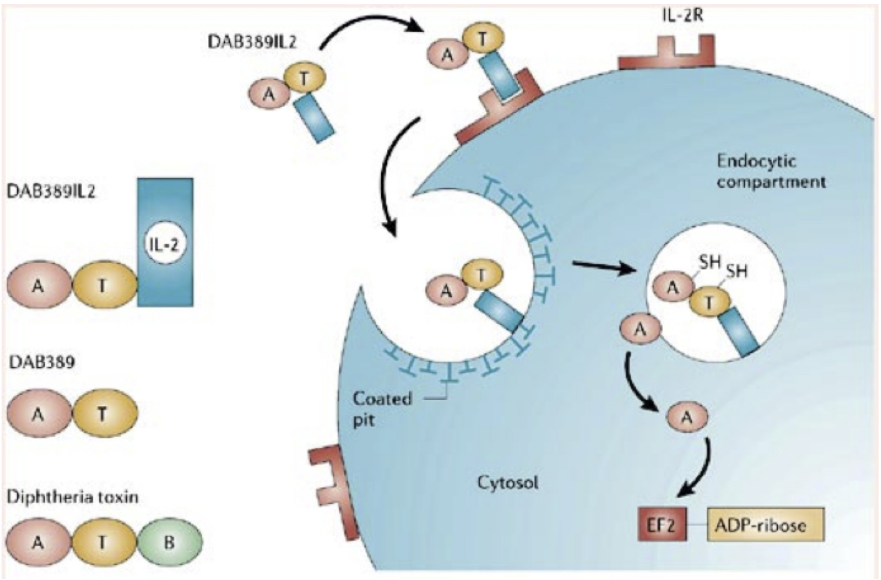

Figure 2.1: An example of immunotoxin. The IT DAB389IL2 is composed of the active and translocating domains of diphtheria toxin (DT) fused to human interleukin 2 (IL2). The immunotoxin binds with cell-surface IL2R and then enters cells. In the cytosol, DT modifies elongation factor 2 (EF2) by adenosine diphosphate (ADP)-ribosylation, which leads to inhibition of protein synthesis and cell death. From (49).

Currently used ITs are derived from various plant and bacterial components, that are then genetically fused or chemically conjugated to ligands that bind to cancer cells (51, 52).

Hence, while in hematologic malignancies the immune system has been damaged, and several cycles of IT therapy can be given without antibody formation, patients with solid tumors and normal immune system may develop antibodies preventing retreatment (53). Non-specific toxicity leading to vascular leak syndrome has also been reported (54). In order to overcome these problems, a novel generation of ITs in which

the cytotoxic moiety is an endogenous protein of human origin, like proapoptotic proteins or RNase, are currently being developed (55).

In this work we focused on Granzyme B. This protein has been demonstrated to be very effective in killing human tumor cell lines that are otherwise resistant against many cytotoxic drugs, and it can easily be recombinantly expressed (56).

2.1.1.1 Granzyme B.

The principal event during the defence against virally infected and malignant cells is the secretion of cytotoxic granules, which contain a variety of granule-associated proteases. Among these proteases, granzymes, structurally related serine proteases, play a significant role. Five different human granzymes are known so far (A, B, H, K and M), each with unique expression patterns and substrate specificities (57). Granzyme B (GB) is one of the most important and well-studied members of the human granzyme family. It is stored as a macro-molecular complex with serglycin and perforin within cytotoxic T lymphocytes (CTLs) and natural killer (NK) cells (58, 59). Its apoptotic activity includes both caspase-dependent and independent pathways and it is based on the proteolytic cleavage of several downstream caspases, predominantly pro-caspase 3, as well as caspase-independent proteins like Bid. These events lead to cytochrome c release and mitochondrial permeability (60, 61, 62, 63). Cytochrome c is a pro-apoptotic factor that, together with HtrA2/OMI (which blocks inhibitors of apoptosis) and endonuclease G (ENDOG; which activates DNA damage), is released after the disruption of the outer mitochondrial membrane by truncated Bid. It then binds to APAF1, which recruits pro-caspase-9 to form the apoptosome. In the apoptosome, pro-caspase-9 is activated and in turn proteolytically activates caspase-3 [1] (Fig. 2.2). Caspase 3, finally, is responsible for DNA fragmentation in apoptosis (65).
GB also disrupts the mitochondrial transmembrane potential through an unknown mechanism and cleaves ICAD (inhibitor of caspase-activated DNase) to free CAD to directly cause DNA fragmentation (64).

[1] Partially activated pro-caspase-3 can also dimerize and activate itself (64).

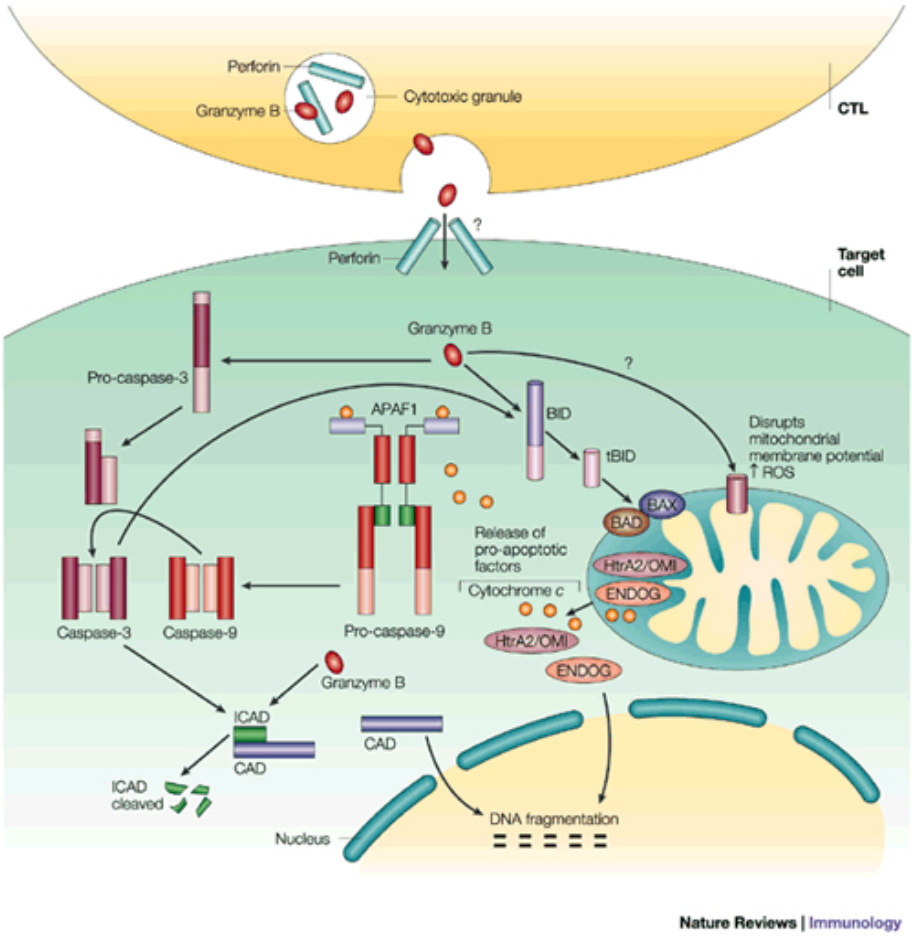

Figure 2.2: Mechanism of action of GB. It enters the cytosol by a mechanism involving perforin and acts through two targets, BID and pro-caspase-3. From (64).

The GB crystal structure has been resolved (66, 67). As in other serine proteases (67), the granzyme B structure is folded into two six-stranded β-barrels connected by three trans-domain segments, and the three catalytic residues, His 44, Asp 88 and Ser 183, are located at the junction of the two β-barrels with the active site cleft perpendicular to this junction (66) (Fig. 2.3).



Figure 2.3: GB structure. The two β-barrels are represented in yellow and orange cartoon, the three catalytic residues in cyan sticks.

2.1.1.2 Serpin B9.

GB-mediated immune responses are in many tumor cells evaded by the cytosolic expression of Serpin B9 (SB9) (68, 69, 70). SB9 is a member of the serpin superfamily, an ever-expanding group of structurally related proteins, which regulate proteases of the classical and innate complement system within vertebrates (71). This 42 kDa protein has been shown to irreversibly inactivate GB (68, 72, 73). This natural GB inhibitor is initially expressed in cytotoxic T lymphocytes and natural killer cells but it is absent from cytotoxic granules. It can protect cells from misdirected GB (73, 74).
It has been proposed that certain tumours up-regulate SerpinB9 as a mechanism of immune evasion (69, 75, 76, 77). This reduces the GB potential for immunotoxin-based therapeutic application (78, 79, 80, 81).
Mammalian serpins are characterized by a common serpin fold, consisting of 350-400 amino acids, three β-sheets and eight or nine α-helical linkers (82). Moreover, a highly flexible peptide loop, known as the reactive centre loop (RCL), holds the reactive centre (Fig. 2.4).

The RCL peptide bond that is attacked by the interacting protease is denoted P1-P1' (84) where P1 is N-terminal and P1' is C-terminal to the site of cleavage. Interaction with the target protease via an ester bond at P1 (85) results in a dramatic conformational change where the RCL is inserted as a middle strand in the central beta-sheet. This change distort and inactivates the interacting protease (83).

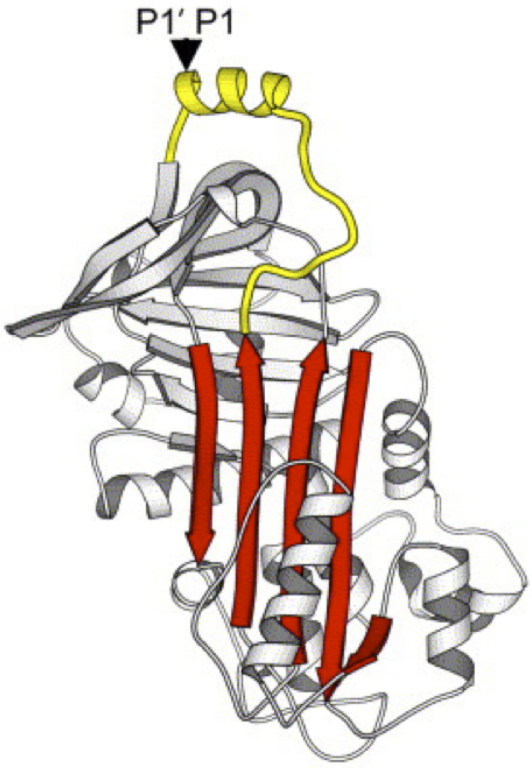

Figure 2.4: Typical serpin fold. RCL is represented in yellow, the central beta-sheet in red. From (83).

2.1.2 Targeted therapy: PARP inhibitors.

A targeted antitumoral therapy that has generated significant interest is the selective cytotoxicity of poly(ADP-ribose) polymerase (PARP) inhibition in tumor cells (86). PARPs, also referred to as diphtheria toxin-like ADP-ribosyltransferases (ARTDs) (87), constitute a multi-gene family of 17 enzymes involved in a number of cellular processes including DNA damage detection and repair, chromatin modification, transcription and cell death pathways. These processes are critical for many physiological and pathophysiological outcomes, including genome maintenance, aging, inflammation, neuronal function and carcinogenesis (88). PARPs' vital role in DNA repair makes them an attractive antitumor target (89).

2.1.2.1 PARPs structure and function.

PARPs are multi-domain proteins, which include: (1) an amino-terminal double zinc-finger DNA binding domain; (2) a central automodification domain, which is responsible for releasing the protein from the DNA after catalysis and (3) a carboxyl-terminal NAD^+ binding catalytic domain (90, 91, 92). The latter is a conserved domain that contains the "PARP signature" motif, a highly conserved sequence (100% conserved in PARP-1 among vertebrates) that forms the active site (91, 93).
PARPs act by catalyzing the polymerization and attachment of ADP-ribose units from NAD^+ onto specific amino acid residues of selected protein substrates (93). In response to stimuli such as a metabolic, chemical, or radiation-induced DNA single-strand break damage, PARPs hydrolyze NAD^+, releasing nicotinamide and one proton, and catalyze the successive transfer of the ADP-ribose moiety to nuclear protein acceptors that might be transiently inactivated. The reaction is initiated by the formation of an ester bond between the amino-acid acceptor (Glu, Asp or COOH-Lys) and the first ADP-ribose. Polymer elongation involves the catalysis of a 2'-1"-glycosidic bond (Fig. 2.5) (94). The up to 500-fold rise in poly(ADP-ribose) synthesis translates DNA damage into intracellular signals that trigger DNA repair and/or cell death pathways (95).

However, some PARP enzymes have mono-ADP-ribose transferase (mART), not ADP-ribose polymer synthase, activity (96). Evidence for this is strong for ARTD8 (PARP14) and ARTD10 (PARP10) but uncertain for other family members, due to the lack of known substrates to study the functions of these enzymes (97).

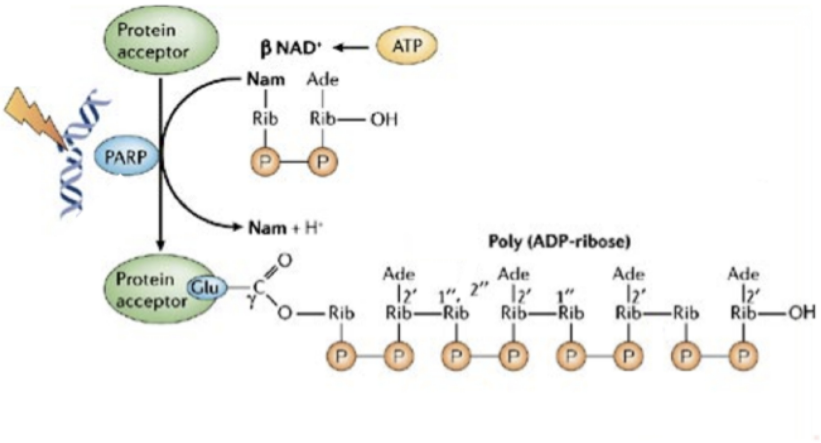

Figure 2.5: Poly-ADP-ribose formation by PARPs. Adapted from (94).

Before the PARP14 and PARP10 characterization, only extracellular enzymes and bacterial toxins were thought to mono-ADP-ribosylate protein targets (98).

In this thesis one of the proteins studied is PARP10, hence in the following section I will describe it more in detail.

2.1.2.2 PARP10.

PARP10 was first identified as a protein interacting with c-MYC, a proto-oncoprotein deregulated in the majority of human cancers that functions as a transcriptional regulator and controls cell proliferation, differentiation, and apoptosis (99).
It is a ubiquitously expressed protein that shuttles between nuclear and cytoplasmic compartments (98) and has auto-ADP-ribosylation activity (100).
On the basis of homology to PARP1 (the best characterized member of the PARP family), PARP10 was expected to function as a poly-ADP-ribose polymerase, whereas it was then identified as a mART with no polymerase activity (34). A comparison of the PARP10 catalytic domain with PARP1 shows that core structural elements such as the NAD^+-binding residues are conserved, but the general domain organization is significantly different (Fig. 2.6) (98).

In particular, the PARP1 H887 and Y919 are involved in NAD^+ binding and conserved in PARP10 and other PARPs. On the contrary, the catalytic glutamate that

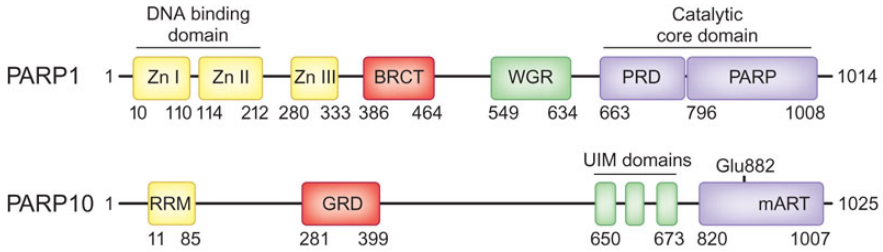

Figure 2.6: Comparison between PARP1 and PARP10 structural organization. PARP1 contains three zinc-finger domains (Zn I, II and III), a breast cancer suppressor protein-1 domain (BRCT), a WGR domain (named after a conserved sequence motif), a PARP regulatory domain (PRD) and the catalytic PARP domain. PARP10 consists of a RNA recognition motif (RRM), a glycine-rich region (GRD) and three ubiquitin-interaction motifs (UIM), in addition to its catalytic mono-ADP-ribosyltransferase (mART) domain. From (98).

defines the enzymatic polymerase activity of PARP1 (E988) is missing in PARP10 as well as in PARP12, and is replaced by an isoleucine (Fig. 2.7) (34).

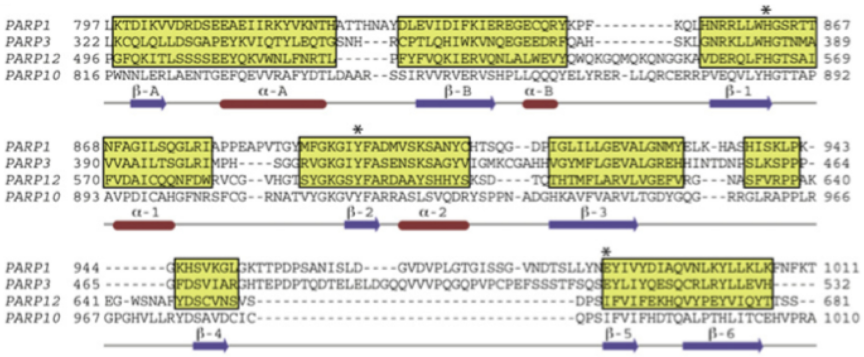

Figure 2.7: Structure-based sequence alignment of PARP1, PARP3, PARP10, and PARP12 catalytic domains. Yellow shadings indicate structurally conserved regions. Asterisks indicate the residues of the conserved "HYE" triad. PARP12 and PARP10 catalytic glutamates are replaced by an isoleucin. From (34).

Thus, an alternative substrate-assisted catalytic mechanism was proposed, in which the acidic target residue of the substrate functionally substitutes for the catalytic glutamate to transfer ADP-ribose. This mechanism explains the fact that PARP10 can only mono-ADP ribosylate its protein substrate (34).

GSK3β. Glycogen synthase kinase-3 (GSK-3), a protein-serine kinase implicated in cell differentiation, phosphorylates several regulatory proteins that are activated by dephosphorylation in response to hormones or growth factors (101). There are two mammalian GSK isoforms encoded by different genes and called GSK3α and GSK3β, which share 85% of sequence identity (102). GSK3β was first identified as a critical mediator in glycogen metabolism and insulin signaling, but it is now known to be an important component of several pathways involved in protein synthesis, cell proliferation and differentiation (103). GSK3β also participates in neoplastic transformation and tumor development. Its role in tumorigenesis and cancer progression remains controversial; it may function as a "tumor suppressor" for certain types of tumors, but promotes growth and development for some others. GSK3β also mediates drug sensitivity/resistance in cancer chemotherapy (104).

Human GSK3β is a 46-47 kDA protein consisting of 433 amino acids. The structure consists of the typical two-domain kinase fold (105) with a β-strand domain at the N-terminal end and a α-helical domain at a C-terminal end. At the interface of the α-helical and β-strand domain, a ATP-binding site is bordered by a glycine-rich loop and a hinge domain (Fig. 2.8) (106).

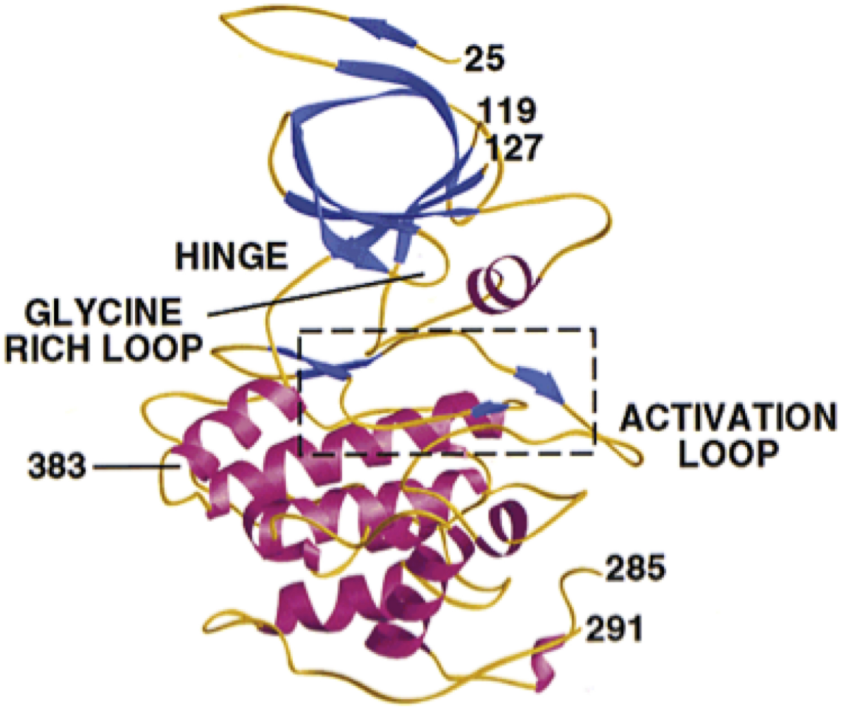

Figure 2.8: GSK3β structure. From (106).

Phosphorylation of Tyr216 increases the catalytic activity, even if it is not strictly required for substrate phosphorylation, while GSK3β phosphorylation at Ser9 in N-terminal region leads to inhibition of its kinase activity (107).
GSK3β has been already found to interact with PARP1 in the regulation of oxidative cell response (108) and with PARP5a (tankyrase) during mitotic progression (109). Understanding the structural details of the GSK3β-PARP10 interaction may be therefore important to develop new therapeutic agents.

2.2 Neurotrophins in neurodegenerative disorders.

Neurotrophins constitute a family of closely related proteins that were identified initially as survival factors for sensory and sympathetic neurons. They promote and regulate the survival of neurons in the vertebrate nervous system and control their development and function (110, 111).
Alterations in neurotrophin levels have been implicated in neurodegenerative disorders such as Huntington's (112) and Alzheimer's (113) disease.
Because of their crucial role for the regulation of synaptic connections and synapse structure, neurotransmitter release and synaptic plasticity, neurotrophins have been studied as possible therapeutic strategies to reduce neuronal degeneration (114, 115, 116, 117, 118, 119, 120).

Neurotrophins family includes the Nerve Growth Factor (NGF), the Brain-Derived Neurotrophic Factor (BDNF), the Neurotrophin-3 (NT-3) and the Neurotrophin-4/5 (NT-4/5) (121, 122). The actions of neurotrophins depend on two different transmembrane-receptor signalling systems (123): the tropomyosin kinase receptors (Trk) and the p75 neurotrophin receptor (p75NTR), a member of the Tumor Necrosis Factor (TNF) receptor superfamily (122, 124).

All neurotrophins bind with similar affinity to p75NTR and selectively to specific Trk receptors, each of them activates several signal transduction pathways (125). NGF binds preferentially to TrkA, BDNF and NT4 to TrkB and NT3 to TrkC (Fig. 2.9). The affinity can be regulated by receptor dimerization, structural modifications or association with the p75NTR (126). The p75 receptor can bind to each neurotrophin,

act as a co-receptor for Trk receptors, increase the affinity of TrkA for NGF and enhance its specificity for cognate neurotrophins (127).

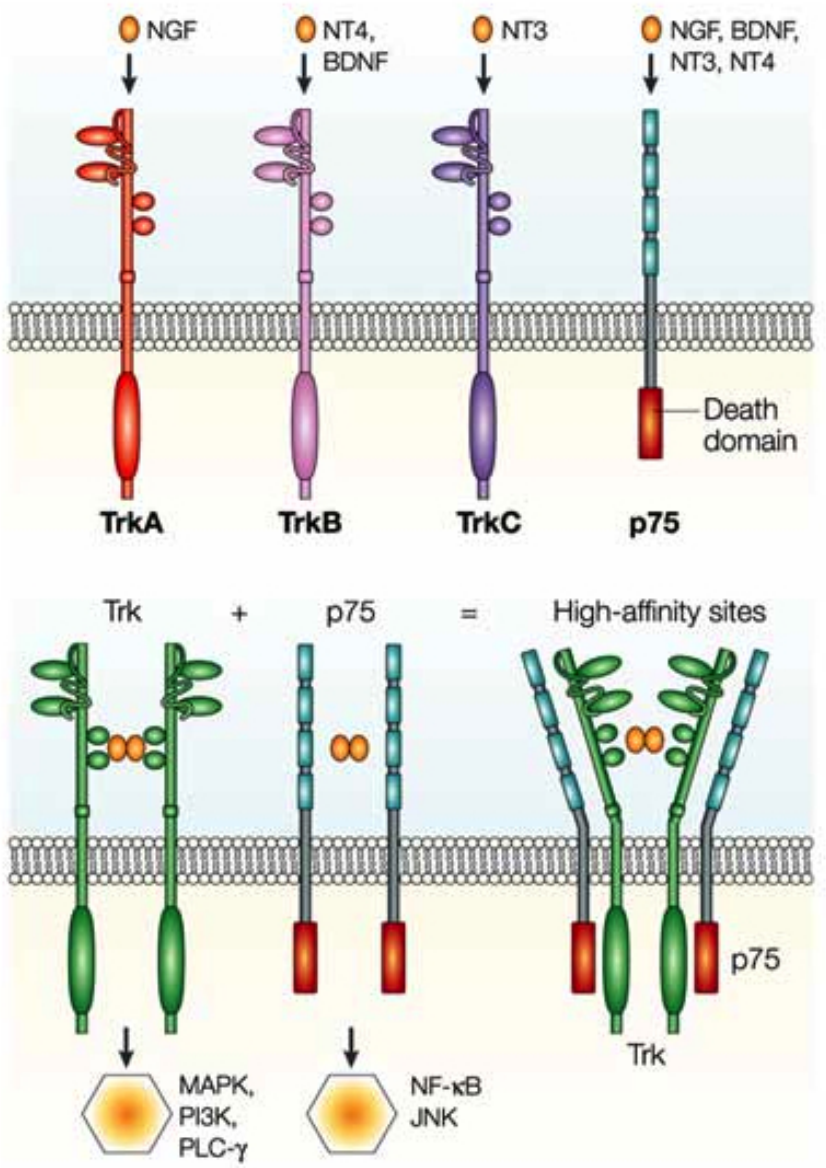

Figure 2.9: Neurotrophin receptors. Neurotrophins binding to two classes of receptors: the low affinity receptor p75NTR and specifically to members of the Trk receptors family. (b) Diagram of the receptors subdomains. Interactions between Trk and p75 receptors can lead to changes in the binding affinity for neurotrophins. From (128).

In the following paragraph I will describe further in detail NGF, the most charac-

terized neurotrophin.

2.2.1 The Nerve Growth factor

The Nerve Growth Factor (NGF) (129) plays a central role in the growth, survival and maintenance of peripheral and central nervous tissues of both embryos and adults (131). In addition, it has been characterized as a signaling molecule for nerve cells and also for the neuroendocrine-immune systems (132, 133).
NGF exerts *in vivo* a number of effects on neuronal and non-neuronal cells (134), being able to i) stimulate the development of sympathetic and sensory neurons (135); ii) guide the nerve growth (136); iii) elicit growth and differentiation into mature cells not only on sensory and sympathetic nerve cells but also on several populations in the central nervous system (137).
An important feature of NGF is the retrograde transport. Indeed, NGF released from target cells activates TrkA receptors on axon terminals and triggers activation of PI3K/Akt, MEK/ERK, and PLC signaling pathways. The signal then travels retrogradely along axon to cell body to promote neuronal survival (138). Moreover, NGF triggers TrkA endocytosis through binding to TrkA and TrkA phosporylation, and the endocytic TrkA might serve as the retrograde signal (139). The TrkA downstream signaling molecules activated by NGF could also act as a retrograde signal (140).

NGF's content in the brain decreases with age, particularly in memory-related areas such as the hippocampus (141). Deficits in NGF release and TrkA phosphorylation signaling have been shown to contribute to age-related decline in long-term potentiation (142). On the other hand, NGF was shown *in vivo* to promote the repair of peripheral nerve injury (143).
However, the therapeutic efficiency of NGF has been greatly limited by delivery problems and severe side effects due to a poor target selectivity (144, 145).
A promising strategy involves the use of peptide mimics, i.e. molecules rationally designed by non-natural modifications of existing peptides to have an agonist or competitive antagonist effect, which may be useful in the treatment of several neurodegenerative diseases and nerve injuries (146, 147, 148, 149). Evidences of the possibility to successfully evaluate the stability of these designed complexes with the use of computational tools have already been provided (150).
NGF mimics have been shown to exhibit neurotrophic effects by binding to TrkA

(151, 152, 153, 154, 155, 156, 157). In this respect, the identification of structural determinants of the NGF/TrkA complex (40) has represented a real breakthrough. Indeed, a deep understanding of the molecular basis of NGF/TrkA molecular recognition might provide important insights for drug design and protein inhibitors (128).

2.2.2 The NGF/TrkA structure.

NGF structure was first determined by X-ray crystallography in 1991 (158). It represents the prototype of a large family of neurotrophins and features a compact tertiary structure based on a cluster of 3 cystine disulfides and 2 very extended, but distorted beta-hairpins (159). It binds to p75NTR (see Appendix B) and TrkA receptors as a dimer held together by non-covalent interactions (160).
The complete extracellular domain of TrkA complexed with NGF was also solved (40), revealing a crab-shaped homodimeric TrkA structure (Fig. 2.10).

The NGF/TrkA structure is characterized by:
(i) a leucine-rich repeat (LRR), formed by three LRR units (D2), flanked by two cysteine-rich domains (D1 and D3). Three leucine-rich repeats form the central part of the D1-D3 body. Each repeats is featured by intrastrand hydrogen-bonded turns. Each of the interstrand segments contains a half-turn of a α-helix. The LRR C-terminal contains a 13-residues α-helix beginning and ending with short 3_{10} helices (154-164 and 169-173). Moreover, it is stabilized by a salt bridge (Glu268-Lys133) and a hydrophobic cluster (Trp132, Trp158, Ala155) surrounded by numerous van der Waals contacts;
(ii) Two immunoglobulin-like subdomains termed IgG-C1 and IgG-C2 (D4 and D5 domain, respectively), deemed relevant for NGF binding (161, 162);
(iii) Loop II of NGF, which was reported to have a role for NGF biological activity mediated by TrkA receptor (163);
(iv) Six high-mannose-containing N-linked glycosylation sites (Asn95, Asn121, Asn188, Asn262, Asn281, Asn358), which appear to play a role in the TrkA structural integrity, as located at TrkA interdomain boundaries.

The secondary structure features of the TrkA in the NGF/TrkA complex include:

- a network of parallel β-sheet involving residues L48 to W144;

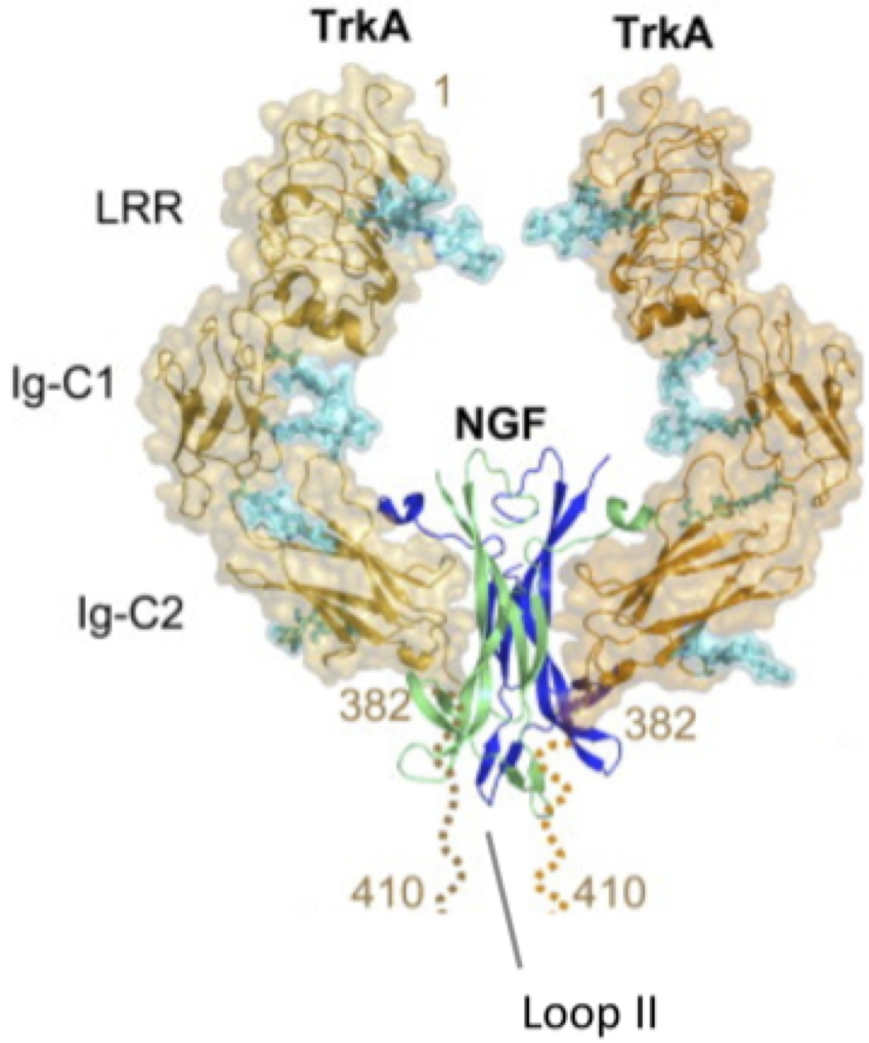

Figure 2.10: Structure of Nerve Growth Factor Complexed with the Extracellular Domain of TrkA. The TrkA ECDs are in orange and shown as ribbons with transparent molecular surface, and NGF is blue-green and only shown as ribbons. N-linked glycosylations observed in the structure are drawn as cyan surfaces. The approximately 32 amino acid connecting peptide linking the base of the Ig-C2 domain of TrkA to the cells surface is drawn as dotted line. Adapted from (40).

- a short 3_{10} helix involving residues C154 to L156 and an α-helix from R157 to E165;

- a second network of β-sheet involving residues W195 to V278;

- a third network of β-sheet involving residues F284 to A376;
- a short 3_{10} helix involving residues H353 to N355.

The structural sections of NGF homodimer include instead:

- a α-helix involving residues P5 to R9;
- a loop-β-sheet involving residues E11 to C58;
- a long parallel β-sheet involving residues W76 to R114.

Zaccaro et al. (164) postulated an allosteric model of TrkA activation by large-scale conformational changes. In this model, TrkA activation is a multistage process involving Trk subdomains, where IgG-C1 subdomain is a hinge region regulating TrkA conformation and activation.
It was already reported in the literature that tyrosine kinase receptors, to which TrkA belongs, undergo conformational changes upon ligand binding (165). However, such a model was argued by Wehrman et al. (40) in the case of TrkA, since the high structural constraint in the NGF/TrkA complex.

In spite of its importance, only very few studies have been focused on the structure and dynamics of NGF and its TrkA receptor (150, 166, 167) and all of them are focused on selected segments of neurotrophins and TrkA.

Through Amber94 (168) -based MD simulations of human NGF-TrkA domain 5 (NGF/TrkA-D5) complex, Settanni et al. (166) found that highly persistent, water-mediated H-bond interactions at the protein-protein interface are relevant for binding. Both direct and water-mediated protein-protein contacts stabilize the complex. Berrera et al. (150) addressed the specificity of NGF N-terminus segment by Amber94-based MD simulations on the NGF/TrkA-D5 complex with or without the NGF N-terminus. The complete complex is stable during the dynamics, but when N-terminus is missing, the two moieties detach. More recently, Amber03 (169)-based MD simulations pointed to the plasticity of the NGF(1-16) and NT-4(1-16) N-terminus peptides (167).

2. MOLECULAR MEDICINE TOPICS INVESTIGATED IN THIS THESIS.

3

Methods

A large amount of data including sequences and structures of several proteins is currently available. Two main computational approaches exist to predict structural determinants and extract functional insights from these data: structural bioinformatics and molecular dynamics.
Bioinformatics uses the statistical analysis of protein sequences and structures to understand their function and to predict structures when sequence information is available. Molecular modeling and atomistic molecular dynamics simulations use the principles from physics and physical chemistry to study the function and folding of proteins.
Both approaches have been used in this thesis. Their fundamental principles are outlined in this Chapter.

3.1 Structural prediction from bioinformatics

3.1.1 Predictions for a single protein

3.1.1.1 Sequence alignment

Sequence alignment aims to match homologous characters that are descended from a common ancestor (170). It can be applied to gene (nucleotides) or protein sequences (aminoacids). The latter application may help to infer the evolutionary relation or also the function of proteins.
The variation of protein sequences during the evolution may affect or not the protein functionality. The mutation probability of each aminoacid into another depends on

their physicochemical similarity. Thus, the alignments can be ranked by assigning a substitution score (171, 172, 173, 174). Evolutionary events such as insertions and deletions in the sequence represent one of the major sources of change at the molecular level (175). These gaps are, however, less frequent events than mutations. They are taken into account in sequence alignment algorithms through the application of gap penalties (176).

Pairwise and multiple sequence alignments. Pairwise sequence alignment (177) consists in maximizing a similarity measure between two sequences. The procedure can be extended to any number of sequences, thereby defining an N-dimensional matrix (178).
The two main classes of pairwise alignments are *global* alignments, which try to maintain a correspondence over the entire length of the two proteins, and *local* alignments, where only the most similar part of the proteins are aligned. Hybrid methods, known as semiglobal or "glocal" (global-local) methods, attempt to find the best possible alignment by creating a map that transforms one sequence into the other while allowing for rearrangement events (179).

Multiple alignments are more informative than pairwise alignments, since their results can be related to various protein structural features (180), but turn out to be much more computational expensive than pairwise alignments (181). After the advent of large genome projects and the consequent explosion of sequence data in public databases, the role of multiple sequence alignments has proved to be very important not only for the identification of conserved motifs or key functional residues in a family of proteins and in phylogenetic inference studies to define the evolutionary relationships between organisms (182), but also to perform comparative multiple sequence analysis at the genome level. This approach, instead of simply transferring annotation from one sequence to another, analyses protein families and their evolution, as well as it detects remote homologs (183). This information confers improved accuracy to most of the computational methods used in proteomics, from gene identification and validation to the determination of the protein 3D structure and the characterization of its molecular and cellular functions (184).

However, while the optimal solution to a pairwise alignment can be found within reasonable time, multiple sequence alignments are significantly more computationally demanding. First of all, it is not straightforward to select the best scoring function according to the phylogeny of the sequences. Hence, they are often assumed to be unrelated. Second, the computational cost of the possible use of dynamic programming (see next paragraph) grows exponentially with the number of sequences.
A commonly used approach is therefore to do *progressive alignment* (185), where multiple alignments are built through the successive construction of pairwise alignments. This permits to deal with a higher number of distantly related sequences. Progressive methods are very dependent on the initial alignment and errors made at any stage of the multiple alignment are propagated through to the final result. Thus, they are more likely to perform well for closely related sequences, which may be seen as a limitation of such methods. On the other hand, progressive methods provide a good compromise between time spent and quality of the resulting alignment, being efficient enough to deal with more than 1000 sequences (182). In addition, many progressive methods that have been developed in the last 15 years that are able to correct the biased selection of the sequences by adding scaling factors weighted on the phylogenetic distance to their scoring function (186, 187, 188, 189) or by including biological information, as in DIALIGN (190) and ClustalW (191). The latter is the most widely used algorithm for multiple sequence alignments, and it has been used in this work. The alignment is achieved via three steps: pairwise alignment (calculating a distance matrix between each pair of sequences), guide-tree generation (constructing a phylogenetic tree that represents the genetic differences among the sequences) and progressive alignment from the nearest to the farther sequences to obtain the multiple alignment (192).

Sequence alignment methods. The three main methods of aligning two protein sequences are (186):
i) dot-matrix methods (193);
ii) dynamic programming (194, 195);
iii) word or k-tuple methods (196, 197).

(i) A dot matrix plot is created by placing one sequence on the horizontal axis and the other on the vertical axis of the matrix. In correspondence of each entry of the

matrix, a point is plotted if the horizontal and vertical elements are identical. If the two sequences are identical, they will be described by a diagonal line in the center of the matrix. The presence of features such as insertions and deletions (causing the sequences to be of different length) leads the diagonal to be broken by empty squares. Mutations and direct or inverted repeats are instead described by multiple lines plotted in a various possibility of configurations (Fig. 3.1).

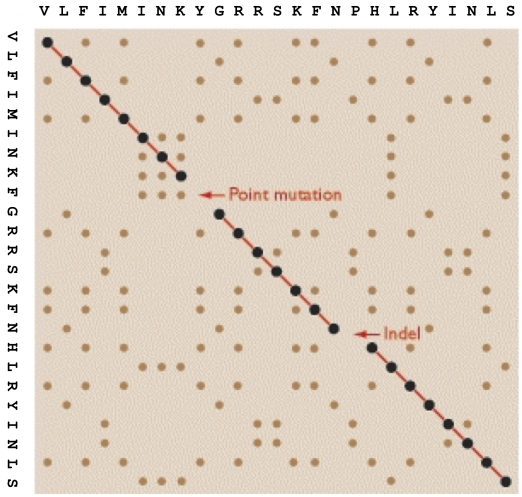

Figure 3.1: The dot matrix technique for sequence alignment. The correct alignment stands out because it forms a diagonal of continuous dots, broken at point mutations and shifting to a different diagonal at indels. Adapted from (198).

(ii) The dynamic programming algorithm uses a substitution matrix to assign match and mismatch scores to aminoacid pairs, and a gap penalty for matching an amino acid in one sequence to a gap in the other. It analyzes this scoring matrix and finds the highest scoring pathway (194) (Fig. 3.2). It consists of three main steps:
1. Initialization: a matrix with *M+1* columns and *N+1* rows is created where M and N correspond to the size of the sequences to be aligned. Assuming that there is no gap opening or gap extension penalty, the first row and first column of the matrix can be initially filled with 0.
2. Matrix fill (scoring): One possible (inefficient) solution of the matrix fill step is based on a procedure that assigns the maximal score $M_{i,j}$ to each entry of the matrix starting from the upper left corner. $M_{i,j}$ depends on the scores associated to the entries to the left, above and diagonal to i and j.

3. Traceback (alignment): The traceback step determines the actual alignment(s) that result a final maximum global alignment score. It takes the current entry and looks to the optimal score neighbor cell among those which could be direct predecessors. There are more alternative solutions each resulting in a maximal global alignment score, but most dynamic programming algorithms will only print out a single solution.

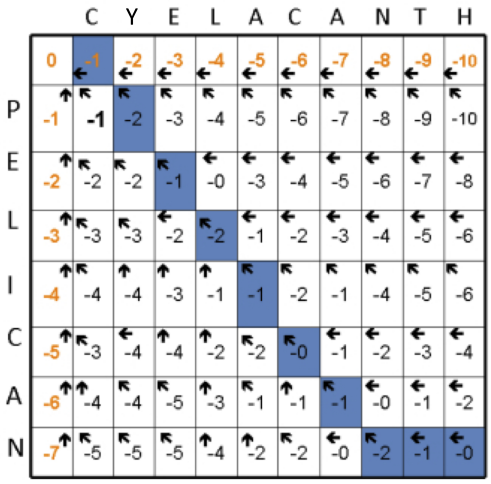

Figure 3.2: An example of dynamic programming for two protein sequences. The cells in the optimum path are showed in blue. Arrowheads are "traceback pointers", indicating which of the three cases were optimal for reaching each cell. Some cells can be reached by two or three different optimal paths of equal score: whenever two or more cases are equally optimal, dynamic programming implementations usually choose one case arbitrarily. Adapted from (199).

The method represents the heart of many well-known programs or their fast approximation (200).

(iii) Word methods, also known as k-tuple methods (201), are heuristic (i.e. experience-based) methods that are not guaranteed to find an optimal alignment solution, but are significantly more efficient than the standard dynamic programming-based methods (i.e. the Needleman-Wunsch algorithm for global alignment and the Smith-Waterman algorithm for local alignment). Word methods compare small sections, with some minimum word length (k-tuples), of each sequence to begin the alignment instead of comparing single letters.

Scoring sequence alignments. The simplest way to score sequence alignments is to count the number of identical residues in two aligned sequences and normalize by sequence length, which yields a percentage identity. However, this approach is too coarse when applied to protein sequences, since it does not take into account two important considerations: i) the different relative frequencies of the twenty amino acids. Certain amino acids, such as alanine and leucine, are abundant, whereas others, such as tryptophan (W) and cysteine (C), are relatively rare. Thus, the conservation of a tryptophan is statistically more significant than conservation of a leucine in a pairwise alignment; ii) the chemical structures and functions of some amino acids, while not identical, may be chemically similar and related (202).
Comparison tables, or matrices, are therefore used to attribute a score to any gene or protein pairwise alignment. For protein sequence alignments, 20 x 20 matrices M are used to evaluate all combinations of aminoacid pairs with a score M(a,b).

Several types of matrices have been proposed for protein sequences. Some are based on a *a priori* estimation of physicochemical similarities. Others are based on the probability of mutation over evolution. For example, PAM (Probability of Acceptable Mutation) matrices (203) are substitution matrices that encode the expected evolutionary change at the amino acid level. Each PAM matrix is designed to compare two sequences which are a specific number of PAM units apart. BLOSUM matrices (204) are instead calculated by considering only blocks of amino acid sequences with small change between them. They are based on a log-odds score calculated for each of the 210 possible substitutions of the 20 standard amino acids, by counting the relative aminoacid frequencies and their substitution probabilities on a database of very conserved regions of protein families, with no alignment gaps.

Large database searches. Sequence database similarity searching is one of the most common computing techniques in modern biology. It allows the identification of sequences homologous to a known sequence, which is used to query large repositories of DNA and protein information. This technique is extremely useful in the study of gene and protein structure, function and evolution. In particular, it is commonly exploited for gene identification, assigning a likely function to a gene or a protein and finding relations among them. An example of application is the identification of a protein

using the sequence of a short peptide fragment identified by biochemistry techniques. However, typical similarity search programs must be especially configured to handle short query sequences, as they are optimized for longer queries. Moreover, aligning a whole sequence database with a query sequence is practically unmanageable with the use of standard sequence alignment algorithms.
Thus, heuristic database similarity search programs such as FASTA (196) and BLAST (197) have been developed, that use 'shortcuts' in order to reduce the time required to build an alignment that has a reasonable chance to be the best one (205).

FASTA. This algorithm (196) is divided into two steps. The first searches for highly similar segments in two sequences. In this search, a word with a specific size is used to identify regions shared by the two sequences with the highest density of identical word matches. The second step is a Smith-Waterman alignment (195), consisting in the search for optimal local alignments, that are discriminated through a scoring system including a set of specified gap penalties. This alignment is centered on the diagonals that correspond to the alignment of the highly similar sequence segments. The region for the Smith-Waterman alignment is bounded by a window size, which limits the number of insertions or deletions one sequence can accumulate with respect to the other sequence in the alignment. Finally, dynamic programming is used to join the segments within this window. FASTA implementation typically use k-tuples of 1 and 2 for protein sequences (206).

BLAST. The BLAST algorithm (197) was developed to perform similarity searches faster than FASTA. As the latter, BLAST also searches for k-tuples, but while FASTA searches for all possible words of the same length, BLAST directly approximates alignments that optimize a measure of local similarity, the so-called maximal segment pair (MSP) score (197) In practice, it achieves its speed by confining the search to the rare words (occurring less frequently in protein sequences) that are most significant (186).

3.1.2 Homology modeling

Introduction. Homology modeling, known also as comparative modeling, is a computational procedure for generating 3D models for proteins of unknown structure and known sequence, based on the alignment with a homologous protein whose structure

is known (template) (207). The proliferation of genome sequencing projects in the last years has been rapidly widening the so-called "sequence gap" between the number of known protein sequences and the number of known protein 3-D structures leading to an increased need for such a kind of technique (208).
Homology modeling is based on the knowledge that the protein structure is determined by its aminoacid sequence, but it is better conserved than the sequence (209). Therefore, proteins with similar sequences, even if distantly related, will fold into similar structures, but also relatively different sequences can share structural similarities (Fig. 3.3). For pairs of distantly related proteins (sequence identity approximately 20%), the regions with the same fold may comprise less than half of each molecule (210, 211). Hence, a "twilight zone" of 20-35% was defined, in which it is not possible to unambiguously distinguish between protein pairs of similar and non-similar structure (212).

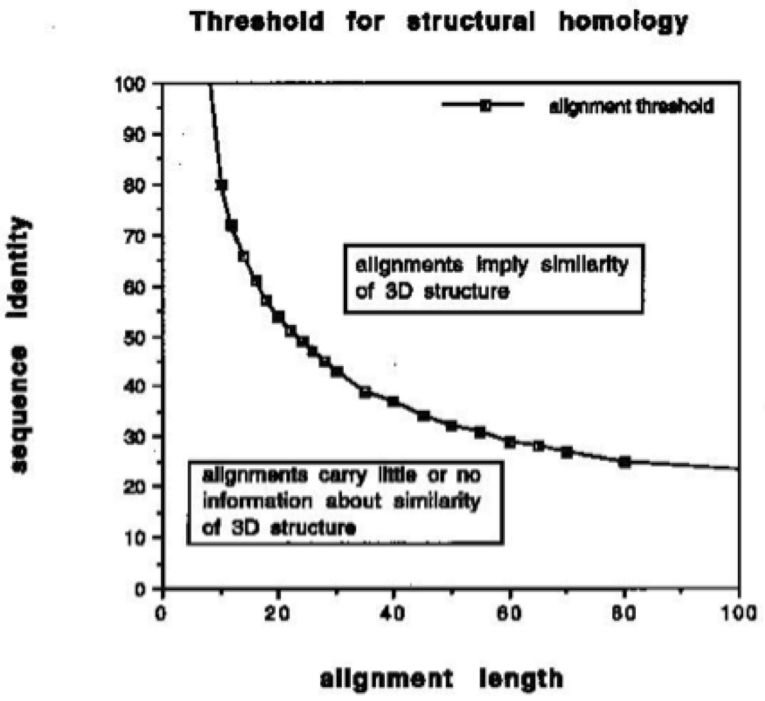

Figure 3.3: Relation between sequence and structure homology. Two proteins share the same 3D-structure when their sequence identity falls in the safe homology zone (upper part of the picture). From (211).

.

The quality of homology modeling at the "twilight zone" may therefore vary widely

(213). These models may however be used to refine NMR structures, to find binding/active sites by 3D motif searching or to predict approximate biochemical function (207). Generally, when the sequence identity exceeds 30%, reliable homology models can instead be constructed (214).

Structural predictions. There are four main steps in constructing models (207): template identification, target-template alignment and correction, model construction (backbone generation, loop modeling, side-chain modeling and model optimization) and model evaluation (Fig. 3.4).

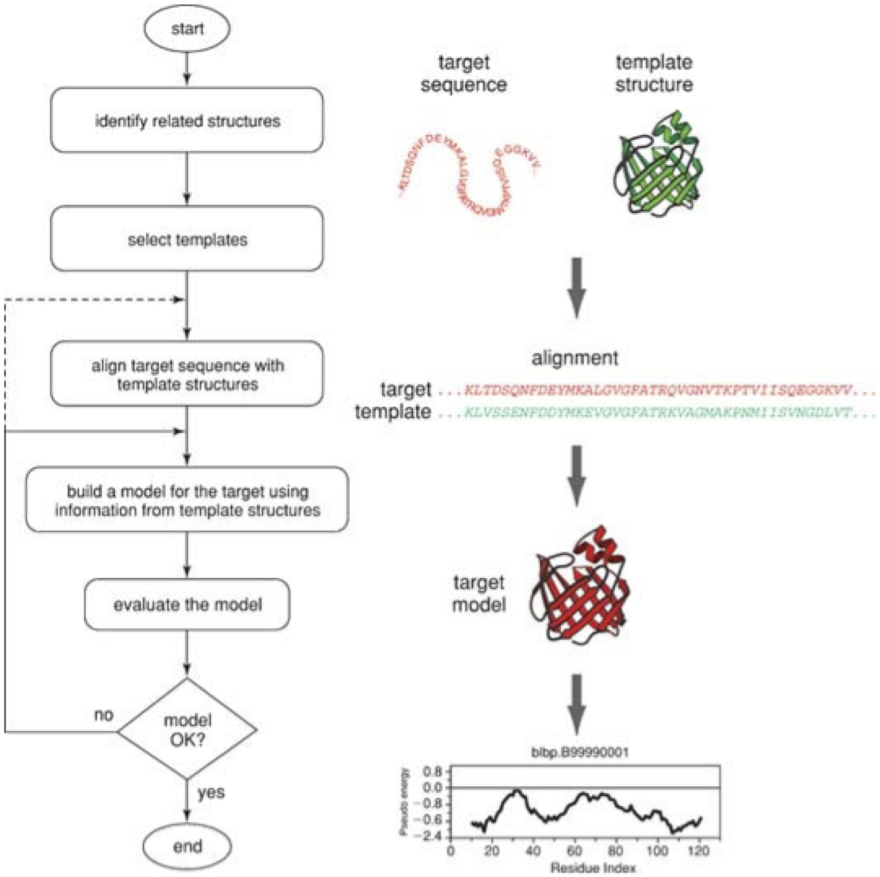

Figure 3.4: Steps in comparative protein structure modeling. From (215).

Template identification. The selection of the template is the initial step. The sequence of the protein with unknown structure is used as a query for search of proteins with known 3D structures in structural databases like the Protein Data Bank (216, 217).

A possible template in the safe percentage identity zone (Fig. 3.3) can be identified with programs such as BLAST (197) or FASTA (218). These programs work by calculating a series of pairwise sequence alignments between the target sequence and each of the database sequences independently, and then providing a set of possible homologs ranked by the scores (219).

Target-template alignment. Once the template is chosen, a pairwise sequence alignment has to be performed using a specific method, because database searching methods do not give an optimal alignment. A specialized alignment method often employed is the ClustalW program (191, 220).
The target-template alignment obtained must be then further improved manually. One technique is to include the sequence of the template and the sequence to be modeled in a multiple sequence alignment with other homologs.
Conserved regions in a multiple sequence alignment of homologous sequences often refer to conserved structural regions, which can help to guide how the template sequence should be aligned to the target sequence. Also, because the template structure reveals what amino acids are located in the core of the protein (normally hydrophobic residues) or on the surface (normally hydrophilic residues), this information can also guide the alignment to correctly align physicochemically similar residues.

Model construction (backbone generation, loop modeling, side-chain modeling and model optimization). Different programs such as MODELLER (221), SWISS PDB Viewer (222), SCWRL (223) and web servers, such as 3D-JIGSAW (224) or SWISS-MODEL (225) are available to build structural models of the proteins based on target-template alignment. For the proteins modeled in this work, the program MODELLER was used (221) because it has been reported to generate homology models characterized by an improved stereochemical quality when compared with other automated programs (226).

MODELLER is based on a technique known as *satisfaction of spatial restraints*, by which a set of geometrical criteria are used to create a probability density function for the location of each atom in the protein. The distance and dihedral angle spatial restraints are combined with stereochemical restraints on bond lengths, bond angles,

dihedral angles and nonbonded contacts based on CHARM22 (227) force field. The probability density function restrains the CA-CA and backbone N-O distances, and backbone and side-chain dihedral angles for different residue types. The generated model violates these restraints as little as possible (228).
Insertions or deletions (gaps) in the alignment are placed out of helices and strands and within loops and turns, but these elements are difficult to predict; moreover, template and target may show quite different loop conformations (229). There are two main approaches to loop modeling: i) Energy based: an energy function is employed, as for true *ab initio* fold prediction. This is the approach exploited by MODELLER; ii) knowledge based: one searches the PDB for known loops with endpoints that match the residues between which the loop has to be inserted, and simply copies the loop conformation (219).
Model refinement is a difficult task that requires an effective sampling strategy as well as an accurate energy functional to guide the search through the conformational space. Homology model refinement is primarily focused on tuning alignment and modeling loops and side chains.
Recent attempts have been made to increase alignment accuracy but they are mainly based on the use of physical chemistry energy information and MD simulations to refine the whole model (230, 231, 232, 233, 234).

Model evaluation. Errors in a homology model may depend on i) the sequence identity between target and template. If it is greater than 90%, the target structure obtained can be compared to a crystallographically determined structure (210); ii) the number of errors in the template. For homology models that have the correct fold, a basic requirement is the stereochemical correctness. Some useful programs for evaluating stereochemistry are PROCHECK (235), PROCHECK-NMR (236), AQUA (236), SQUID (237), WHATCHECK (238). The features of a model that are checked by these programs include bond length, bond angles, side chain torsion angles, the location of main chain conformation in acceptable regions of the Ramachandran map, the planarity of side chain rings and of peptide bonds, a proper environment for hydrophobic and hydrophilic residues, the absence of bad atom-atom contacts and of structural holes (239). Based on these features, scoring functions such as the MolProbity clashscore (240), and the PROCHECK G-factors (236) can help distinguishing between homology

models of higher and lower accuracy (241).
Other evaluation methods assess the environment of each residue with respect to the expected environment as found in the high resolution X-ray structures. These methods are based on 3D profiles and statistical potentials of mean force (242) and are aimed at finding structures/folds compatible with a given sequence. Programs implementing this approach include Verify3D (243) and Anolea (244).
Ultimately, the validation of models comes from experiments such as site-directed mutagenesis, circular dichroism, cross-linking, mass spectrometry, fluorescence-based thermal shift, light scattering, molecular FRET or electron microscopy. Such experimental data can be translated into constraints/restraints and introduced in the modeling protocols thus improving the accuracy of the models. Faster crosschecks are also possible: for instance, with enzymes it is possible to verify the location of important catalytic residues in the active site by comparison with homologous family members (245).
Since 1994, the biennial Critical Assessment of Structure Prediction (CASP) contests (246) have provided an ideal opportunity to evaluate the accuracy of todays many protein structure prediction methods. During each CASP season, about 200 research groups try to predict the structures of more than 100 CASP targets. The target sequences are provided to CASP by structural biology laboratories just before the corresponding structures are solved. The predictions are thus true blind predictions, allowing performance to be measured on realistic test cases. CASP regularly shows that the homology modelling steps described above allow reliable models to be built in many cases, from which a lot of structural and functional insights can be derived (247).

3.1.3 Protein-protein complex structural predictions.

3.1.3.1 Introduction.

Protein-protein docking methods can create models of protein complexes, provided that the conformational changes upon association are restricted to the side chains (22). However, it remains very challenging to account for backbone conformational changes happening upon binding (22). Hence, most algorithms treat proteins as rigid bodies or permit side-chain motion only, in order to reduce the search in the conformational space (248). On the other hand, new promising approaches of flexible refinement,

ensemble docking and explicit inclusion of flexibility during the entire docking process are constantly being developed (249).

3.1.3.2 Multistep procedure in docking.

The general scheme of docking can be divided into four major stages (250):

1. *Preprocessing stage*: the proteins are analyzed to define their conformational space and interaction surfaces. This process simulates the conformational selection model, in which the conformational change in binding is thought to originate primarily from the conformational diversity of the unbound state (251). The analysis of residues important for the protein-protein interaction can be supported by user-provided bioinformatic or experimental data.
2. *Rigid docking*: a set of solutions is generated including at least one near-native structure. This implies the possible occurrence of steric clashes, because proteins in their unbound conformation can collide when placed in their native interacting position.
3. *Refinement*: each candidate is optimized by small backbone and side-chain movements and by rigid-body adjustments. The resulting refined structures have better binding energy and hardly include steric clashes. This stage models an induced fit (252), a process where the initial interaction between a protein and a binding partner induces a conformational change in the protein through a stepwise process (253).
4. *Scoring*: the candidate solutions are scored and ranked according to different parameters such as force-field based potentials, agreement with known binding sites, deformation energy of the flexible proteins, buried surface area and desolvation energy terms. This stage is aimed to identify the near-native solutions among all the best-ranked candidates.

In this work we used HADDOCK (High Ambiguity Driven protein-protein Docking) algorithm (21, 254). Its docking protocol consists of three stages:
i) a rigid-body energy minimization;
ii) a semi-flexible refinement in torsion angle space;
iii) a final refinement in explicit solvent.
After each of these stages, structures are scored and ranked, and the best structures are kept for the next stage (255). In particular:

i) In the initial randomization step, the two molecules, A and B, are separated by 25 $\AA$ and randomly rotated around their center of mass. Then they are allowed to rigidly rotate to minimize the intermolecular energy function. Two cycles of translational and rotational rigid body minimization are subsequently performed, and the two molecules are docked. The best structures in scoring term (tipically 200 structures) will proceed to the second step.
ii) This step consists of a semi-flexible simulated annealing refinement in which the side chains and the backbone at the interface are allowed to move, in order to take into account some possible conformational rearrangements. This semi-flexible simulated annealing (SA) consists of several stages: a) high temperature rigid body search (2000K, 500 steps); b) rigid body SA (cooling step from 2000K to 500K, 500 steps); c) semi-flexible SA with flexible side-chains at the interface (from 1000K to 50K, 1500 steps); d) semi-flexible SA with fully flexible interface, i.e. including both backbone and side-chains (from 500K to 50K, 1500 steps). The resulting structures undergo then steepest descent energy minimization.
iii) The final step consists of a gentle MD simulation refinement in a 8 $\AA$ shell of TIP3P (256) water molecules of the remaining best structures (usually 200). During this stage, the system is first heated to 300K, with flexible side-chains at the interface. 1500 molecular dynamics steps at 300K are performed, with a position restraint only on non-interface heavy atoms. Finally the system is cooled down (1000 steps at 300, 200, 100K) imposing position restraints only to backbone atoms outside the interface.

Finally, the docking solution are clustered (based on pairwise backbone RMSD at the interface) and sorted based on Haddock score. Clusters are analyzed and ranked according to their average interaction energies (sum of electrostatic, van der Waals terms, desolvation energy, Ambiguous Interaction Restraints (AIR) energy terms) and their average buried surface area. The interaction energies are evaluated by full electrostatic and van der Waals energy terms with a 8.5 $\AA$ distance cutoff using the OPLS (257) force field.
To get accurate docking predictions, HADDOCK takes as input also bioinformatics information, biochemical and/or biophysical experimental data, like mutagenesis data or information about chemical shift perturbation upon binding. This information is used to define the distance restraints that will be used along the docking of two up to

6 biomolecules.

The Ambiguous Interaction Restraints (AIR) is a collection of restraints on two set of residues, named active and passive residues, on the interacting surface. Active residues are for example those detected by experiments (mutagenesis, chemical shift perturbation upon complex formation) to make contact within the complex and solvent accessible. Passive residues are their closest solvent accessible neighbors that potentially make contacts. For every active residue, a single AIR restraint is defined between that residue and all active and passive residues on the partner. An explicit AIR energy term is introduced into the calculation with the functional form of a harmonic potential (becoming linear after a given cutoff distance) that depends on an effective distance. The latter is calculated through the following formula:

$$d_{i_{AB}}^{eff} = \left(\sum_{m_i A=1}^{Natoms} \sum_{k=1}^{Nres_B} \sum_{n_k=1}^{Natoms} \frac{1}{d_{mn_k}^6} \right)^{-\frac{1}{6}} \tag{3.1}$$

where m and n run over all the atoms of a given residue and k over the sum of active and passive residues for a given protein. An upper limit to the effective distance (typically 2 $\mathring{A}$) is enforced by Haddock (21). If this limit is exceeded, the AIR energy becomes positive and the active residue experiences an attractive force towards the active and passive residues of the partner molecule. If not, the restraint is satisfied and the AIR energy and attractive force are zero for that restraint. Since many atom-atom distances inversely contribute to the effective distance, an AIR restraint is typically satisfied if a residue comes within 4-5 $\mathring{A}$ of any active or passive residue of the partner molecule (258).

It is also possible to randomly remove a fraction of the AIRs, if the definition of interacting residues is not clear. In this way, bad restraints could be discarded, allowing for better docking solutions.

3.1.3.3 Model evaluation.

The *PISA method* (259) calculates the ΔG as the solvation free energy gain upon formation of the predicted complex, in kcal/mol. The value is calculated as difference in total solvation energies of isolated and interfacing structures. This prediction method is validated by experimental measurements of solvation energies (259), and the values obtained should be compared with the contribution from the interface interactions, i.e.

mainly salt bridges (0.9-1.25 kcal/mol) (260), hydrogen bonds (0.6-1.5 kcal/mol) (261) and at a lower occurrence disulphide bridges (2-8 kcal/mol) (262).

Alanine scanning methodology is aimed at measuring the effect of the deletion of an amino acid side chain beyond the Cβ carbon atom on the affinity of a protein-protein complex, thus allowing the identification of putative hot spots. Unfortunately, despite significant advances in molecular biology, experimental alanine scanning still represents a large effort that cannot be applied easily to a large scale screening of protein-protein interfaces. Hence, a variety of computational approaches has been developed to detect hot spots in modeled or experimentally determined protein-protein complexes for which no experimental mutagenesis data is available (263).
In particular we used the Baker's approach (28, 264). This method first defines the interface residues as (i) residues that has at least one atom within a 4 $\mathring{A}$ radius sphere of an atom belonging to the protein partner, or (ii) residues that becomes significantly buried upon complex formation, as measured by an increase in the number of Cβ atoms within a sphere with a radius of 8 $\mathring{A}$ around the Cβ atom of the residue of interest. The program then replaces each of the interface residues individually with alanine residues, and computes the effect of this mutation on the binding free energy of the complex, allowing the identification of potential hot spots. This estimation relies on a simple physical model. It uses all heavy atoms and polar hydrogens to represent proteins and proposes a free energy function to linearly combine terms such as Lennard-Jones potentials to describe atomic packing interactions, an orientation-dependent hydrogen bond potential derived from high-resolution protein structures (265), Coulomb electrostatics, and an implicit solvation model (266).
This model is validated by using mutational data for protein complexes from the Alanine Scanning Energetics database (ASEdb) (267), a searchable database of binding energy changes resulting from mutations of protein side-chain to alanine.

However, for the reliability of protein-protein docking results it is also crucial to have an experimental validation. In most studies, models are selected and validated using a variety of well-described methods able to provide some information on the identity of interacting residues. These techniques include (268, 269): i) site-directed mutagenesis (270); ii) fluorescence-resonance energy transfer (FRET) (271), used to

determine the distance between labeled groups of interacting proteins; iii) chemical crosslinking with subsequent mass spectrometric identication of the crosslinked peptides after proteolytic digestion, a hybrid technique which appears especially well-suited to capture information on residues involved in transient complexes (272, 273); iv) cross-saturation coupled with NMR TROSY (transverse relaxation optimized spectroscopy) experiments (274), which can be used for the identification of the interfaces of large protein-protein complexes in solution (275, 276).

3.2 Physics-based methods: Molecular dynamics and free energy calculations.

3.2.1 Introduction

Molecular dynamics (MD) is an important tool to understand the physical basis of the structure and function of biological molecules. This computational method calculates the time dependent behavior of a molecular system and is now routinely used to investigate the structure, dynamics and thermodynamics of biological molecules and their complexes. Indeed, the early view of proteins as relatively rigid structures has been replaced by a dynamic model in which the internal motions and resulting conformational changes play an essential role in their function (277).
In biophysics, the most widely used MD technique is the so-called classical MD. Such method describes the atoms as point-particles by neglecting their internal structure, and in particular the role of the electrons belonging to them. Such an approximation results a completely adequate tool to describe the motion of any biologically relevant system as far as chemical reactions are not involved in the mechanism we want to investigate.

Classical MD is based on Newton's equation of motion, i.e. the atomic dynamic trajectories are extracted by integrating the Newton's second law:

$$\vec{F}_i = m_i \vec{a}_i \text{ with } \vec{F}_i = \frac{\delta V(\vec{r})}{\delta \vec{r}_i} \tag{3.2}$$

where $\vec{r}_i$ is the position of the *i-th* atom, $\vec{F}_i$ the force acting on it, and V is the potential energy of the system which is a function of the atom positions. The potential

V that defines the interactions among the particles is described by a *force field*, i.e. a particular functional form dependent on several empirical parameters obtained by fitting experimental data or data coming from theoretical calculations based on high-level (and computationally very expensive) quantum theories.
The ergodic hypothesis guarantees that thermodynamics information on the system can be estimated as time averages, which equal the statistical ensemble average:

$$\frac{1}{T}\int_0^\infty A(t)dt = \langle A \rangle \tag{3.3}$$

In fact, for any system for which the ergodic hypothesis holds, we can say that such system will eventually pass through all possible states if one allows the system to evolve in time indefinitely. Therefore, in performing a molecular dynamics simulation one needs to generate a longer and longer trajectory so that this equality is better and better satisfied. In this case, experimentally relevant information concerning structural, dynamic and thermodynamic properties may then be calculated using a feasible amount of computer resources (278).

In the typical MD algorithm cycle, the first step is to assign the initial velocities and positions to each particle of the system. In the MD simulations used for the 'production run', the initial coordinates for a MD simulation usually comes from preliminary steps such as an energy minimization, where a set of coordinates that minimizes the total energy of the system and satisfies any given constraint is found; when the temperature is given, the initial velocities are usually randomly generated at the beginning of the simulation only constrained to satisfy the Maxwell distribution at that temperature. The force acting on each particle is then calculated by using the chosen force field and the position of all the other atoms. Any modern force field is formed by two parts that are very different from the computational point of view: the short-range interaction and the long-range interaction terms. The short-range interaction terms represent those physical terms that becomes negligible when the particles come far apart. Therefore, to save computational time, these terms are usually evaluated only for particles that are within a certain fixed cutoff radius. The Newton's equation of motion is then solved by some integration method and finally the configuration of the system is updated. The entire cycle just described is indefinitely repeated to produce the MD trajectory, with the only difference in the initial step above where now the initial coordinates and

velocities will be the updated coordinates and velocities obtained in the last step.

In this work, the NAMD simulation package (279) has been used. In the following sections, the main MD integration algorithms and the functional form of the force field are described.

3.2.2 Integration of the equations of motions.

MD simulations consist in the numerical, step-by-step solution of the classical equations of motion (Eq. 3.1). The integration time step Δt is chosen at the beginning of the simulation and it remains unchanged during the run. The time step must be small enough to allow describing the fastest motions of the system: $\Delta t \leq 0.5$ fs is normally used when bond stretching have to be described. The fastest motions in a simulation of biological systems will invariably involve the lightest hydrogen atoms (280). By constraining these bonds to their equilibrium lengths using the LINCS algorithm (281), a larger $\Delta t \leq 1.5/2$ fs can be employed while still retaining a good accuracy.
The most used algorithms for the integration of the equations of motions are the so-called Verlet (282) and Leap-Frog (283) algorithms.
To derive the Verlet algorithm we start with the Taylor series of the position of a particle around the time t and with the time step Δt

$$\vec{r}(t+\Delta t) = \vec{r}(t) + \frac{d\vec{r}(t)}{dt}\Delta t + \frac{d^2\vec{r}(t)}{dt^2}\frac{\Delta t^2}{2!} + \frac{d^3\vec{r}(t)}{dt^3}\frac{\Delta t^3}{3!} + O(\Delta t)^4 \tag{3.4}$$

The first order derivative of *r(t)* with respect to t is equal to the velocity. Similarly the second order derivative equals to the acceleration, which can be substituted by the force by using Newton's second law. This reduces eq. 3.4 to

$$\vec{r}(t+\Delta t) = \vec{r}(t) + \vec{v}(t)\Delta t + \frac{\vec{F}(t)}{2m}\delta t^2 + \frac{d^3\vec{r}(t)}{dt^3}\frac{\Delta t^3}{3!} + O(\Delta t^4) \tag{3.5}$$

and similarly we have

$$\vec{r}(t-\Delta t) = \vec{r}(t) - \vec{v}(t)\Delta t + \frac{\vec{F}(t)}{2m}\delta t^2 - \frac{d^3\vec{r}(t)}{dt^3}\frac{\Delta t^3}{3!} + O(\Delta t^4) \tag{3.6}$$

Summing these two last equations and neglecting the higher order terms gives us

$$\vec{r}(t+\Delta t) \approx 2\vec{r}(t) - \vec{r}(t-\Delta t) + \frac{\vec{F}(t)}{m}\Delta t^2 \tag{3.7}$$

where $\vec{r}(t+\Delta t)$ is the new position of the particle, $\vec{r}(t)$ its present position, $\vec{r}(t-\Delta t)$ the position on the previous time step, $\vec{F}(t)$ the resulting force on the particle, and m the mass of the particle. The Verlet algorithm does not use the velocity of a particle to compute its new position.

The leapfrog algorithm can be derived from the Verlet algorithm. Leapfrog evaluates the velocities at half-integer time steps, which are artificial points in time, exactly half way between two time steps. Hence, to compute the new position, the algorithm uses the velocities

$$\vec{v}(t - \frac{\Delta t}{2}) \equiv \frac{\vec{r}(t) - \vec{r}(t - \Delta t)}{\Delta t} \tag{3.8}$$

and

$$\vec{v}(t + \frac{\Delta t}{2}) \equiv \frac{\vec{r}(t - \Delta t) - \vec{r}(t)}{\Delta t} \tag{3.9}$$

Thus, assuming that the initial positions $\vec{r}(t)$ and velocity $\vec{v}(t - \frac{\Delta t}{2})$ are known, the steps to solve the equations of motions using the leapfrog algorithm are:
i) calculate $\vec{F}(t)$;
ii) calculate the new velocity $\vec{v}(t + \frac{\Delta t}{2})$ by exploiting $\vec{v}(t - \frac{\Delta t}{2})$ and $\vec{F}(t)$;
iii) calculate the new position $\vec{r}(t + \frac{\Delta t}{2})$ by using eq. 3.9 with $\vec{r}(t)$ and $\vec{r}(v + \frac{\Delta t}{2})$;
iv) repeat these steps until a stop criterium. The leapfrog method is however algebraically equivalent to Verlet's algorithm, and, consequently, it gives identical trajectories (278). However, there are some advantages in using the leapfrog scheme. The algorithm is very stable and faster than Verlet; in addition, because of the symmetric way in which it is defined, it is time reversible. This feature guarantees the conservation of conserved quantities such as energy and angular momentum.

3.2.3 Force field.

A force field is a mathematical expression describing the dependence of the potential energy *U* of a system on the coordinates of its particles. It consists of an analytical form of the interatomic potential energy, *U(r1 , r2 , . . . , rN)* and a set of parameters entering into this form. The parameters are typically obtained either from *ab initio* or semi-empirical quantum mechanical calculations or by fitting to experimental data

such as neutron, X-ray and electron diffraction, NMR, infrared, Raman and neutron spectroscopy, etc (284). In this work the AMBER parm99SB force field was used (285). The AMBER functional form is:

$$U = \sum_{bonds} \frac{1}{2} k_b (r - r_0)^2 + \sum_{angles} \frac{1}{2} k_a (\theta - \theta_0)^2 + \sum_{torsions} \frac{V_n}{2} [1 + \cos(n\phi - \delta)] \quad (3.10)$$

$$+ \sum_{impropers} V_{imp} + \sum LJ4\epsilon_{ij} \left(\frac{\sigma_{ij}^{12}}{r_{ij}^{12}} - \frac{\sigma_{ij}^{6}}{r_{ij}^{6}}\right) + \sum_{elec} \frac{q_i q_j}{r_{ij}} \quad (3.11)$$

where the first four terms refer to intramolecular or local contributions to the total energy (bond stretching, angle bending, and dihedral and improper torsions), and the last two terms serve to describe the Van der Waals interactions (in this case by means of a 12-6 Lennard-Jones potential) and the Coulombic interactions.

Bond stretching and angle bending are very often represented with a simple harmonic function also in other force fields. A dihedral or torsional term is also required in any molecule containing more than four atoms in a row. While angle bending, and bond stretching, are often not relevant high frequency motions that can be replaced by a rigid approximation, torsional motions are typically hundreds of times less stiff than bond stretching motions and they are necessary to reproduce the major conformational changes due to rotations about bonds and to correctly describe the local structure of a macromolecule (286). Torsional energy is usually represented by a cosine function. The parameters are usually derived from *ab initio* calculations and then refined using experimental data such as vibrational spectra. The improper torsion term is needed to ensure the planarity of some particular groups, such as sp2 hybridized carbons in carbonyl groups or in aromatic rings, and describes the positive contribution to the energy of those out-of-plane motions.

The last two terms describe the non-bonded intermolecular van der Waals and electrostatic interactions. Van der Waals interactions between two atoms arise from the balance between repulsive and attractive forces. Physically, the repulsion origins from the overlap of the electron clouds of the atoms, while the interactions between induced dipoles result in an attractive component that can be modeled with a r^{-6} term.

Concerning electrostatic interactions, while the molecular electronic density can be obtained with a high accuracy by means of high-level quantum-mechanical calculations,

the problem of reducing such density to a manageable description to be used in a MD simulation is not trivial. The usual choice is to assign a partial atomic charge to each nucleus and use Coulomb's law to compute their contribution to the total energy. In particular, the dielectric constant is assumed equal to 1 (vacuum value), and the restrained electrostatic potential model (287) is used to define partial atomic point charges. In this model, charged are assigned to the atom-centered points so as to fit the electrostatic potential derived from quantum chemistry calculation for a set of small representative molecules.

3.2.4 Periodic boundary conditions.

In most cases simulations are not aimed to study isolated (in vacuo) particles, but the bulk properties of a liquid or solid system. In particular, due to the importance of water in connection with the properties of biological macromolecules, they are commonly simulated in aqueous solution rather than in the gas phase.
However, standard simulations track only a small number of particles in order not to slow down the computation. As a result, in these conditions most molecules would be near the edge of the system, resulting in unrealistic surface effects.
For this reason, the imposition of some boundary conditions is required. With the introduction of periodic boundary conditions (PBC), the simulation box is surrounded by an infinite number of replicas of itself (Fig. 3.5). Only the N atoms inside the main cell are considered explicitly, but as soon as one of the atoms leaves the cell, an image particle enters from the opposite side to replace it.
Typically a cutoff radius is used for Lennard-Jones interactions of the order of 10 Å. If the potential range is not too long (the cutoff radius do not exceed half of the box size), the *minimum image convention* is adopted in the calculation of the Lennard-Jones term. This means that each individual particle in the simulation interacts only with one atom, the closest one, among all its images.
In the case of macromolecules simulated in solution, their interaction with periodic images would be an artifact and should be avoided. Thus, the length of each box vector must exceed the length of the macromolecule in the direction of that edge plus twice the cutoff radius.

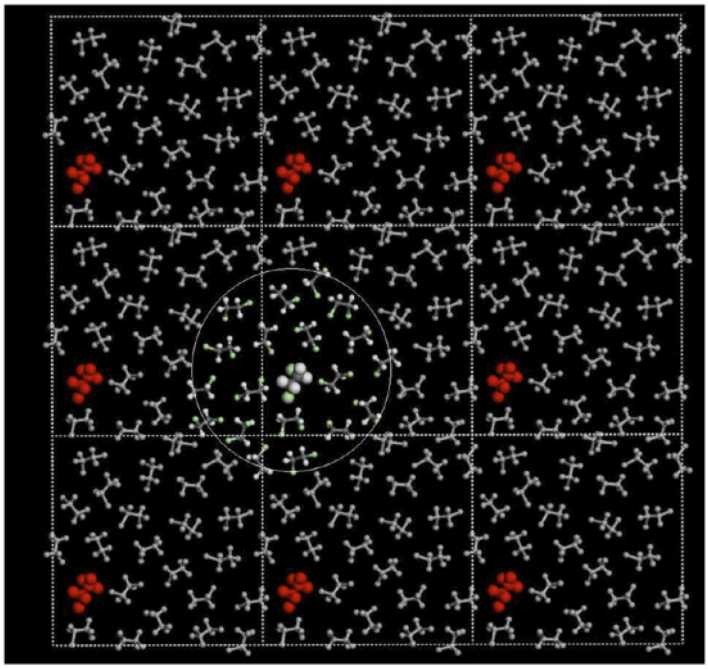

Figure 3.5: Schematic representation of periodic boundary conditions in 2D. The figure shows a selected molecule in the primary cell together with 8 replicas. The colored molecules represent those being at a distance from the reference molecule smaller than the cutoff radius. Typically the cutoff is applied to full molecules (for example using the distance between their center of masses as reference) or neutral groups, in order to minimize truncation effects. (Taken from (286)).

3.2.5 Neighbors list.

The computation of the short-range van der Waals terms in eq. 3.11 involves, in principle, a large number of pairwise calculations. However, only a relatively small number of pairs are within the cutoff radius for the van der Waals interactions. Therefore, in order to speed up the simulation, a neighbors list is commonly employed.

Namely, in a system composed of N atoms, for each atom i the distances with the remaining *N-1* atoms are calculated. The short-range non-bonded contribution to the total force acting on i is due only to the atoms within a sphere of a certain radius R_c from the atom i. A list of all of the atoms that are at a distance smaller than R_c+Δ_{list} is created for each atom i, where Δ_{list} is a value chosen by the user (typically $\sim$ 1-2 Å). At each time step, the search for the atoms situated at a distance $\leq R_c$ of a given atom

is done only over the atoms included in its neighbor list. The list is updated regularly but only after a certain number of step (e.g. 20) to take into account the atoms that enter or leave the sphere.

3.2.6 Long range interactions.

An interaction between to atoms is defined long-range when the corresponding force falls off to zero no faster than r^{-d}, where d is the dimensionality of the system (d=3 in common biological cases). In recent years, a number of methods have been introduced which allows the inclusion of long-range electrostatic interactions in molecular dynamics simulations. For simulations of proteins and enzymes in a crystalline state, the Ewald summation is considered to be the best treatment for long-range electrostatic interactions (288). Some approximated but numerically faster variations of the Ewald summation method include the particle-mesh Ewald method (289) that is employed in the NAMD program.

3.2.6.1 Ewald sum.

The Ewald sum (290) is a method to sum up electrostatic interactions in an infinite lattice, or, alternatively to compute electrostatic interactions under periodic boundary conditions. The potential energy can be written as

$$V = \frac{1}{2}\sum(\frac{1}{4\pi\epsilon_0}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{q_i q_j}{r_{ij}, n}) \tag{3.12}$$

where q_i and q_j are the charges, n is the index of the periodic box and r_{ij} is the distance between the charges i and j in the *n-th* box. Unfortunately, this sum converges slowly with n and therefore a very large number of images is required to achieve a reliable estimation of the potential energy. The idea behind this method is to:
i) surround each ion with a Gaussian charge distribution of opposite sign and equal magnitude to screen the interactions so that they are short-ranged (and the sum of interactions is now absolutely convergent):

$$\rho_i^G(r) = q_i(\frac{\alpha}{\pi})^{\frac{3}{2}}\exp(-\alpha|r_i + nL|^2) \tag{3.13}$$

where α, the width of the Gaussian distribution, is an arbitrary parameter that can be adjusted to optimize the convergence rate;
ii) add a neutralizing charge distribution so that the overall potential becomes identical to the original one;
iii) calculate the contribution to the electrostatic potential of this this neutralizing distribution in the reciprocal space where it is absolutely convergent and therefore both the terms, the short range one in ii) and this one, will be absolutely convergent and suitable for a numerical treatment.

3.2.6.2 Particle Mesh Ewald.

The particle mesh Ewald (PME) approach allows a fast calculation of the electrostatic energy in PBC, using the profound advantages of the fast Fourier transform (FFT) for calculating discrete Fourier transforms. When a fixed cutoff is applied to the direct Ewald sum, the number of terms needed in the reciprocal sum is *O(N)*. Since each such term is nominally of order N to calculate, the reciprocal sum is $O(N^2)$. Similarly, the discrete Fourier transform of N coefficients is an order N^2 calculation. However, the FFT performs this task in $O(N)\log N$ operations. All the particle mesh-based approaches reduce the calculation of the reciprocal sum to a sum over coefficients of the discrete Fourier transform of a gridded charge distribution, which is then accelerated to an $O(N)\log N$ calculation using the three-dimensional FFT (291). Therefore, the PME algorithm is particularly suited to simulate Coulombic electrostatic interactions of periodic systems with a large number N of atoms.

3.2.7 Temperature and pressure control.

MD simulations can be carried out in different thermodynamic ensembles, characterized by the control of certain thermodynamic quantities. The simplest is the microcanonical ensemble (NVE), in which the number of particles, the volume and the total energy are conserved. However, the microcanonical ensemble does not correspond to the ensemble in typical experimental conditions. A more meaningful description of the physical system is achieved by simulating the system at varying energy but with pressure and/or temperature that are kept constant. These are called canonical (NVT) or isothermal-isobaric (NPT) ensembles. In these simulations, thermostat and barostat algorithms

(the main ones are described below) are required to control the temperature and pressure of the system during the MD run. In this work the Nose-Hoover thermostat and barostat are adopted (292, 293).

3.2.7.1 Berendsen thermostat.

Experimentally, to maintain the temperature, the system is coupled to an external heat bath with fixed temperature T_0. In the Berendsen approach, the thermostat behavior is mimicked by rescaling the velocities of the particles at each step, such that the rate of change of temperature is proportional to the difference in temperature:

$$\frac{dT(t)}{dt} = \frac{1}{\tau}(T_0 - T(t)) \tag{3.14}$$

where τ represents the coupling parameter which determines how tightly the bath and the system are coupled together. This method gives an exponential decay of the system towards the desired temperature. The change in temperature between successive time steps is

$$\Delta(t) = \frac{\delta(t)}{\tau}(T_0 - T(t)) \tag{3.15}$$

Thus, the scaling factor for the velocities is

$$\lambda^2 = 1 + \frac{\delta(t)}{\tau}\frac{T_0}{T(t - \frac{\delta(t)}{2})} - 1 \tag{3.16}$$

The $\frac{\delta(t)}{2}$ term is due to the fact that the called leap-frog algorithm is used for the time integration.

In practice, τ is used as an empirical parameter to adjust the strength of the coupling. Its value has to be chosen with care. In the limit $\tau \to \infty$ the Berendsen thermostat is inactive and the run is sampling a microcanonical ensemble. The temperature fluctuations will grow until they reach the appropriate value of a microcanonical ensemble. However, they will never reach the appropriate value for a canonical ensemble. On the other hand, too small values of τ will cause unrealistically low temperature fluctuations. If τ is chosen the same as the timestep $\delta(t)$, the Berendsen thermostat is nothing else than the simple velocity scaling. Values of ~ 0.1 ps are typically used in MD simulations of condensed-phase systems.

3.2.7.2 Nose-Hoover temperature coupling.

The extended system method introduced by Nose and subsequently developed by Hoover (292, 293) turns out to be more efficient for sampling a correct canonical ensemble than the Berendsen thermostat, which is instead suitably used for relaxing a system to a target temperature.
The idea of the Nose-Hoover thermostat is to modify the equation of motions to include a non-Newtonian term in order to maintain the total kinetic energy constant. This term is a friction factor used to control particle velocities. This friction factor is actually the scaled velocity, $v(\xi)$, of an additional and dimensionless degree of freedom ξ. This degree of freedom has an associated "mass", which effectively determines the strength of the thermostat. The equations of motions obeyed by this additional degree of freedom guarantee that the original degrees of freedom sample a canonical ensemble.

3.2.7.3 Langevin thermostat.

In the Langevin thermostat (294), an additional friction force β and a small random thermal noise term η_i extracted from a Gaussian distribution are added to the velocity of each system particle:

$$\vec{r_i}(t) = \frac{\vec{F_i}}{m_i} = \beta\vec{r_i}(t) + \eta_i(t) \tag{3.17}$$

These two factors are balanced in order to give on average a constant temperature.

3.2.7.4 Berendsen barostat.

The Berendsen algorithm to control the pressure of the system is similar to the Berendsen temperature coupling. In this case, the system pressure is set toward a desired value by changing the dimensions of the simulation cell size during the simulation. An extra term is added to the equations of motion that effects a pressure change:

$$(\frac{dp}{dt})_{bath} = \frac{p_0 - p}{\tau_p} \tag{3.18}$$

where τ_p is the time constant for the coupling. A simple proportional coordinate scaling, concomitant with volume scaling, minimizes local disturbances.

3.2.7.5 Nose-Hoover / Langevin piston.

The pressure control schemes used in this thesis is the Nose-Hoover / Langevin piston implemented in NAMD (279). This method should be combined with a method of temperature control, such as Langevin dynamics, in order to simulate the NPT ensemble. Such a barostat is a combination of the Nose-Hoover constant pressure method (295). where an extra degree of freedom corresponding to a piston is added into the equations of motions, and the piston fluctuation control is implemented by using a Langevin dynamics as in (296).

The complete set of equations of motion used by NAMD when both the barostat and the thermostat are enabled then are:

$$r' = \frac{p}{m} + \beta r$$

$$p' = F - \beta p + gp + \eta$$

$$V = 3V'\beta$$

$$\beta' = \frac{3V}{W}(P - P_0) - g_\epsilon \beta + \frac{\eta_\epsilon}{W}$$

$$W = 3N\tau^2 kT$$

$$\langle \eta^2 \rangle = \frac{2mgkT}{h}$$

$$\tau = oscillationperiod$$

$$\langle \eta_\epsilon^2 \rangle = \frac{2Wg_\epsilon kT}{h} \tag{3.19}$$

where β is the friction coefficient (thermostat degree of freedom), W is the mass of piston, τ is the barostat time constant, η is noise on atoms, and η_ϵ is the noise on the piston. The piston represents an additional degree of freedom, which undergoes a Langevin like dynamics, in analogy with the motion of each particle.

3.2.8 Free energy calculations.

There are problems in biophysics for which a straightforward MD simulation can usually not provide solutions due to the limitation in the length of the trajectories accumulated. For example, a system could remain localized in a certain region during the entire MD simulation because the free energy barriers to overcome are too high and the Boltzmann probability that the system in a certain time can have sufficient energy for that is simply too small that also a single event is not observed in the time of the simulation. Another example is the case of protein folding where the folding time of the most of the proteins is of the order of seconds, while direct simulation of these systems can access the microseconds time scale at most. For this reason, the problem of the correct sampling is of great importance and several free energy methods have been developed to efficiently sample events that involve the passage through high free energy barriers.

These free energy methods can be divided into:

i) methods aimed at reconstructing the probability distribution or the free energy as a function of one or more coordinates. Examples are thermodynamic integration (297, 298), where the free energy difference between a certain reference state and the state of interest is measured by changing infinitely slowly the thermodynamic parameters characterizing the system so that at each stage along the path the system is in equilibrium; umbrella sampling (299) which applies an artificial biasing potential to enhance sampling of unfavorable regions along a reaction coordinate; parallel tempering, where elevated different temperatures are used to overcome barriers on several replicas of the system that are simulated simultaneously;

ii) methods directly aimed at accelerating rare events and constructing reactive trajectories, such as steered MD (300) based on the addition of external forces which reduce the energy barriers (301).

The metadynamics method (302) provides in many cases a unified framework for computing free energies and for accelerating rare events (303, 304, 305, 306). This approach consists in a continuous addition of an history-dependent potential energy to the energy potential of the system, that forces the dynamics to explore previously not visited conformations. Metadynamics requires the preliminary identification of a set of collective variables (CVs) that are functions of those atomic coordinates that describe the

activated process of interest.
The forces acting on each atom are therefore corrected by the addition of a history-dependent external contribution, and this correction can be simply obtained as the derivative of the history-dependent potential energy V_t, with respect to the atom coordinate. V_t is constructed as the sum of a set of Gaussian functions centered along the trajectory

$$V(s,t) = \sum_{t'<t} wexp\left(- \frac{s - s_{t'}^2}{2\delta^2} \right) \tag{3.20}$$

where s is some CV, s_t is the value of s along the trajectory at at time t and the sum run over all the times before t when a Gaussian deposition has been performed.
In this way, the system is discouraged from revisiting configurations that have already been sampled.
After exploring all conformations defined within the CVs space, the free energy profile does not change anymore and can be easily reconstructed as the opposite of the sum of all Gaussians.
One of the more widely used versions of metadynamics is commonly referred to as "well-tempered" metadynamics (307). In this approach, the sampling of the CVs is performed at the enhanced temperature $T+\Delta T$ and the local strength of the Gaussian functions that are used to bias the sampling are reduced by an automatic rescaling.
This is achieved by using a different expression for the bias potential:

$$V(s,t) = k_B \Delta T \ln \left(1 + \frac{W}{\tau} \frac{N(s,t)}{k_B \Delta T} \right) \tag{3.21}$$

where *N(s,t)* is the histogram of the s variables collected during the simulation, ΔT an input parameter with the dimension of a temperature, W an input parameter with the dimension of an energy, and τ the Gaussian deposition rate. This guarantees that the bias potential converges in a single simulation in the time limit, and does not oscillate around the free energy profile solution. Indeed, as the bias deposition rate decreases as *1/t*, the dynamics of all the microscopic variables becomes progressively closer to the thermodynamic equilibrium as the simulation proceeds (309).

4

Design of Human Granzyme B Variants resistant to Serpin B9.

Human Granzyme B (hGB) is a serine protease involved in immune-mediated apoptosis. Its cytotoxicity makes it potentially applicable in cancer therapy. However, the effectiveness of hGB can be hampered by the cytosolic expression of a natural protein inhibitor, human Serpin B9 (hSB9). Here we used computational approaches to identify hGB mutations that can affect its binding to hSB9 without significantly decreasing its catalytic efficiency. Alanine-scanning calculations allowed us to identify residues of hGB important for the interaction with hSB9. Some variants were selected, and molecular dynamics simulations on the mutated hGB in complex with hSB9 in aqueous solution were carried out by Prof. Barth's group (RWTH Aachen) to investigate the effect of these variants on the stability of the complex. The R28K, R201A and R201K mutants significantly destabilize the interaction of the protein with hSB9. Consistently, all of these variants also retained their activity in presence of the Serpin B9 inhibitor in subsequent in vitro assays of wild-type and mutated hGB. In particular, the activity of R201K hGB with and without Serpin B9 is very similar to that of the wild-type protein. Hence, R201K hGB emerges as a promising species for anti-tumoral therapy applications.

4.1 Introduction.

Several strategies for anticancer therapy are based on human serine protease Granzyme B (hGB) (31). hGB is the most important and well-studied member of the granzyme family, secreted during the defence against virally infected and malignant cells. Each

member exhibits unique expression patterns and substrate specificities (57). hGB manages to kill tumor cell lines that are otherwise resistant against many cytotoxic drugs. Its apoptotic activity is based on the proteolytic cleavage of several downstream caspases and caspase substrates as well as caspase-independent proteins like Bid (310, 311). This leads to cytochrome c release and mitochondrial permeability (60, 312, 313, 314). These processes are crucial for apoptosis of, e.g., tumor cells (315, 316).
hGB-based therapeutic applications, being based on an endogenous protein of human origin, have the great advantage of avoiding neutralizing antibody formation. This prevents patient retreatment, as reported for constructs derived from various plant and bacterial components (53). However, they are associated with a few drawbacks as well (56, 317). One of the most severe of them is the inhibition of hGBs enzymatic activity by human Serpin B9 (hSB9) (68, 72, 73, 318). Indeed, the hGB·hSB9 complex has practically zero catalytic activity (69, 72). At the molecular level, hSB9 inhibits hGB by forming a reversible Michaelis complex, followed by a 1:1 covalent complex (319, 320): a covalent bond is formed between the side chain oxygen atom of hGBs *S183* (italics are used to indicate hGB residues) and the carbonyl carbon of hSB9's P1 residue (321).

Identifying hGB variants able to bind more weakly with hSB9 and thus retaining (at least in part) protease activity may therefore dramatically increase the potential of hGB for antitumoral applications. In order to rationally design such mutants, investigators would benefit greatly from experimental structural information on the full-length hGB·hSB9 noncovalent or covalent complexes. Unfortunately, such information is currently lacking.
Here we have addressed this issue by predicting the structure of the hGB·hSB9's Michaelis complex and by testing the predictions accuracy by measuring in vitro the enzymatic activity of the wt hGB protein and its mutants in the presence of hSB9. Our calculations were based on the structurally similar rat trypsin·*M. sexta* serpin complex (319) and hGB (67) X-ray structures. We then identified key interactions at the interface of the two proteins using the Robetta (28, 264) webserver, based on our structural prediction.
Among the wild-type (wt) hGB residues involved in such interactions, we selected residues which were not located in hGB's active site. These are expected not to affect hGB's catalytic activity. Next, we designed several mutants (including *R28K*

hGB·hSB9, *R201A* hGB·hSB9 and *R201K* hGB·hSB9) that could decrease the affinity of the complex using computational alanine scanning mutagenesis (CASM).
CASM is a convenient and computationally fast strategy to investigate the determinants of protein complex interactions (322), although it provides very approximate results. In particular, CASM does not include hydration at the protein/protein interface (264) in spite of the fact that water molecules may play a very important role for complex stability. Furthermore, CASM does not allow to investigate the impact of mutations (other than alanine) on the structural determinants. Finally, CASM does not allow to take into account fully the flexibility properties of the complex. Therefore, we decided to perform molecular dynamics (MD) simulations of wt and mutated hGB·hSB9 complexes in water solution, which allow us to successfully overcome the limitations of the CASM approach albeit at the price of a greater computational cost. Free energy calculations, which could provide quantitatively changes of the mutants relatively to that of the wild-type, are computationally more expensive and provide quantities which can only be related indirectly to enzymatic catalysis.
Those calculations allowed us to suggest that the three mutants above are destabilized to the greatest extent among the mutants investigated here. On the other hand, the others turned out to be almost as stable as the wt hGB·hSB9 complex. Consistently, the measured enzymatic activity of the three mutants above was similar to that of wt hGB. These mutants might be therefore promising for exploitation in therapeutic applications, depending upon further assessment of a putative intrinsic immunogenicity due to the mutation itself.

4.2 Methods.

Molecular Modeling. We predicted the structure of human hGB·hSB9 complex by homology modeling based on the rat trypsin·*M. sexta* serpin B1 complex X-ray structure (PDB ID: 1K9O) (319). These are suitable templates for hGB and hSB9. The sequence identity (SI) between rat trypsin and hGB is 36%. The two proteins have similar length (Fig. 4.1a), and they are functionally and structurally related (310, 323, 324). We thus fitted the backbone of hGB structure with that of rat trypsin in complex with *M. sexta* serpin B1 (Fig. A.1 in Appendix).

4. DESIGN OF HUMAN GRANZYME B VARIANTS RESISTANT TO SERPIN B9.

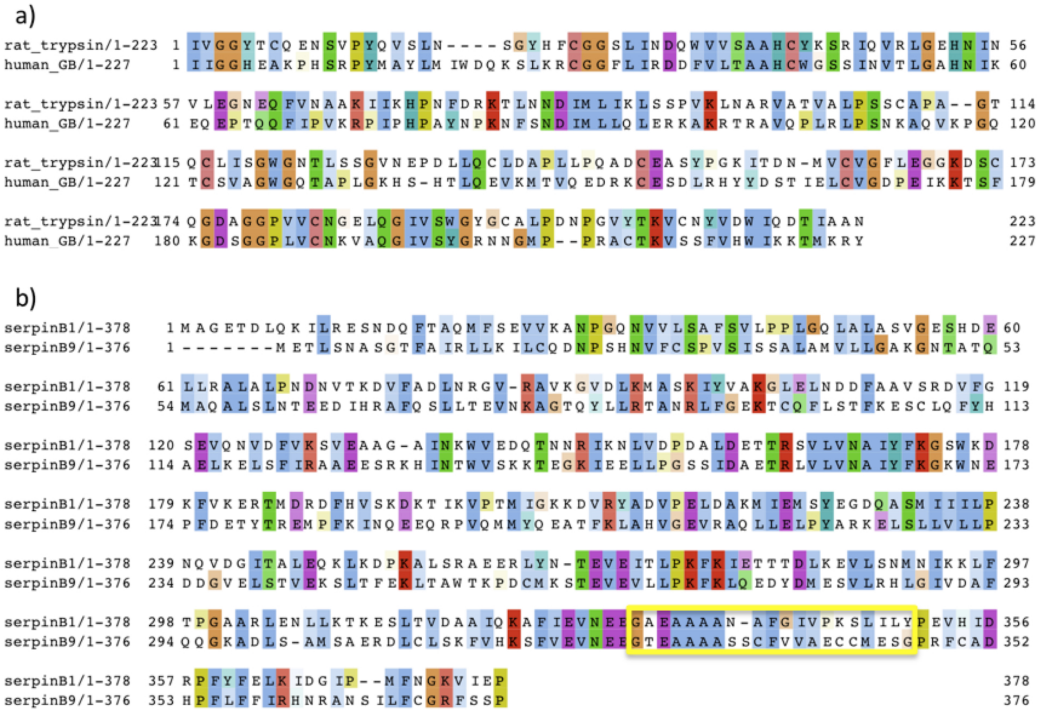

Figure 4.1: Sequence alignments. Sequence alignments between rat trypsin and hGB (a) and of *M. sexta* Serpin B1 and hSB9 (b), as obtained by using ClustalW (http://www.ebi.ac.uk/clustalw). The hSB9's active site residues are inserted in the yellow box.

Two X-ray structures of hGB are available. These are the unbound form (66) (PDB ID: 1FQ3) solved at a resolution of 3.6 Å, and the form bound to a 5-residues peptide (67), (PDB ID 1IAU) solved at 2 *Å* resolution. The latter is considered in this work. The SI between *M. sexta* serpin B1 and hSB9 is 27% (Fig. 4.1b). Furthermore, serpin B1's active site is strongly conserved across the serpin family (325), including SB9 (Fig. 4.1b). Because the structure of hSB9 has not yet been solved, we built its structure using homology modeling based upon the X-ray structure of the rat trypsin·*M.sexta* serpin B1 complex (Fig. A.1 in Appendix), using MODELLERv7 software (221). The top-scoring structure was identified based upon its stereochemical quality, checked using the PROCHECK software (326) as well as its DOPE (Discrete Optimized Protein Energy) score (327), a statistical potential optimized for model assessment implemented in the MODELLER software. The final RMSD between hSB9 and Serpin B1 and that between human hGB and rat trypsin were both 0.7 Å.

The complex was inserted in a water box of edges 108×81×119 $Å^3$ including ∼29,000 water molecules and 7 Cl- counterions. Periodic boundary conditions were applied. It first underwent 5000 steps minimization and then 30 ns of molecular dynamics (MD) at 300 K. The AMBER ff99SB (285), TIP3P (256) and Aaqvist (328) force fields were used for the protein, water and counterions, respectively. The timestep was 2 fs. The SHAKE algorithm (329) was applied to all chemical bonds. The particle mesh Ewald method was applied to evaluate the long-range electrostatic interactions (330). Constant temperature and pressure were achieved by coupling the system with a Langevin thermostat (294) and a Nose-Hoover-Langevin barostat (296).

In the last few years, a number of methods have been developed for the identification of hot spots in protein-protein interactions. They are represented on the one hand by rigorous treatments based on physical effective energy functions such as free energy perturbation. Because these methods that require a sufficient statistical sampling, they are often extremely time and computationally consuming (331, 332). On the other hand, these methods may be based on empirical treatments. These include Molecular Mechanics/Poisson-Boltzmann surface area (MM-PBSA) approaches (333), which use a simplified continuum solvent method. Computational alanine scanning coupled with MD simulations, used here, might represent a good compromise between speed and accuracy.

Mutants putatively destabilizing the complex and not affecting the active site structures were therefore identified using Baker's alanine scanning procedure (28, 264). Residues predicted to significantly ($\geq$ 1 kcal/mol) decrease the binding free energy upon mutation to alanine, thus destabilizing the interface, are defined as hot spots. *K27* hGB, *R28* hGB and *R201* hGB hotspots were identified here (Table A.1, Fig A.1 in Appendix). Putative variants affecting the interactions between hSB9 and the hotspots were built using the SWISS PDB Viewer software (222). These are: *K27A* hGB·hSB9, *R28A* hGB·hSB9, *R201A* hGB·hSB9, *R28E* hGB·hSB9, *R28K* hGB·hSB9, *R201E* hGB·hSB9, and *R201K* hGB·hSB9, as well as the double mutant *R28A-R201A* hGB·hSB9. No clashes were present in the predicted structures.

These variants were inserted in water boxes and counterions were added, similarly to what was done for wild-type (wt) hGB·hSB9 (Table A.2 in Appendix). 50 ns MD simulations were carried out for the variants as for wt hGB·hSB9. A cluster analysis (334) was performed to identify representative structures of the conformational ensemble of the wt hGB·hSB9 complex and to identify the variants which were stable during the dynamics: *R28A* hGB·hSB9, *R201E* hGB·hSB9 and *K27A* hGB·hSB9 (see Results Section). One single representative from each of those ensembles covered from 52% to 83% of the conformational space simulated. Thus, some of the structural properties presented here are calculated for that representative. The NAMD 2.7 package was used for all MD simulations (279).

Experimental methods. The cloning of the plasmid encoding the sequence of the hGB fusion protein as well as the expression in HEK293T cells has been described before (79). The transfection was done with 1 μg DNA according to the manufacturer's using RotiFect (Roth). The protein was secreted into the supernatant (335) and purified via Immobilized Metal-ion Affinity Chromatography (IMAC). The cleared supernatant was supplemented with 10 mM imidazole and loaded to an XK16/20 column (Amersham / GE Healthcare) containing 8 ml Sepharose 6 Fast Flow resin (Clontech / Takara). The incubation, washing and elution buffers used, were as described previously (335). The eluted protein was re-buffered into 20 mM Tris, pH 7.4, 50 mM NaCl, concentrated, aliquoted and stored at -80C. For the generation of an active protein with a free N-terminus (63, 336), Enterokinase was added to the protein (0.02 U/μg with 2 mM $CaCl_2$ for 16 h incubation at 23 °C prior to use. The protein concentration was calculated after SDS-PAGE analysis and Coomassie staining using AIDA Image Analyzer Software (Raytest Isotopenmessgeräte).
For the expression of recombinant hSB9 the *E.coli* expression strain BL21 (DE3) was used during a 4 L fermentation in synthetic medium. The bacteria were harvested 24 h after induction with Isopropyl Thiogalactoside (IPTG). After centrifugation the bacterial pellet was re-suspended in lysis buffer (50 mM NaH_2PO_4 , pH 8.0, 500 mM NaCl, 0.5 mM DTT) and sonicated six times for 60 s, 70%, 9 cycles. The supernatant was supplemented with 10 mM imidazol and was purified via IMAC as described above. After a second IMAC, the protein was re-buffered into 20 mM Tris, pH 7.4, 1 mM DTT and further purified via anion exchange chromatography (Q-Sepharose XL, 1 ml, GE

Healthcare) with a salt gradient of 0.05 - 1 M NaCl. For higher purity, size exclusion chromatography (SEC) was done as the last step, using a Superdex 75 (GE Healthcare) column in 20 mM Tris, pH 7.5, 50 mM NaCl, 1 mM DTT.
The complex formation between recombinant hSB9 and the hGB has been described before (68, 72). 600 ng wild-type or mutated hGB fusion protein were incubated with or without hSB9 (3-fold excess) for 1 hour at 37 °C under reducing conditions in 20 mM Tris, pH 7.4, 50 mM NaCl and 1 mM DTT. The remaining activity was detected by cleavage of 200 μM substrate Ac-IETD-pNA (Calbiochem/Merck) imitating the cleavage site of Pro-caspase 3. The reaction was monitored in a microplate reader in triplets at 37 °C with an absorbance of 405 nm. The velocity of the activity was calculated from the linear slope of the reaction during the first 10 minutes.

4.3 Results and discussion

We discuss here the structural features of wild-type (wt) and mutants of hGB in complex with hSB9 in water solution, as obtained by 50 ns of MD in explicit solvent.

wt hGB·hSB9 complex. The Root Mean Square Displacement of the wt complex with respect to the initial structure (RMSD hereafter) fluctuates along the MD trajectory around 3.2 ± 0.2 $\AA$ for the last 20 ns of the MD run (Table 4.1, Fig. 4.2). This suggests that the structure is equilibrated after 30 ns. The following analysis is performed therefore for the last 20 ns. The distance between hGB's and SB9's centers of mass (DCOM hereafter) is 49 ± 0.4 Å, and the solvent-accessible surface area (SASA) (337) is 29,680 ± 371 Å^2 (Table 4.1, Fig. 4.2). The protein-protein contact surface as obtained by cluster analysis (see Method section) involves 4 salt bridges, 2 hydrogen bonds and 3 hydrophobic contacts (Table 4.2, Fig. 4.3). They are significantly different from those of the template (Table 4.2, Table A.3 in Appendix, Fig. 4.3).

Design of mutated wt hGB·hSB9 complexes. We identified mutations that might affect the stability of the complex, without altering hGB enzymatic activity.

Complex	RMSD(Å)	DCOM(Å)	SASA(Å^2)
wt hSB·hSB9	3.2 ± 0.2 3.4	48.9 ± 0.4 49.1	29680.5 ± 371.0 29736.7
R28A hSB·hSB9	3.6 ± 0.3 3.6	49.4 ± 0.5 50.3	29715.1 ± 296.1 29789.3
R28E hSB·hSB9	5.5 ± 0.5 4.9	49.6 ± 0.6 49.2	30156.7 ± 298.3 29713.1
R28K hSB·hSB9	= 4.9	= 52.6	= 30395.8
R201A hSB·hSB9	= 3.9	= 47.0	= 31053.1
R201E hSB·hSB9	2.7 ± 0.1 2.6	49.3 ± 0.2 49.7	29395.5 ± 251.6 29505.0
R201K hSB·hSB9	= 3.4	= 49.1	= 29736.7
R28A-R201A hSB·hSB9	5.7 ± 0.6 6.8	48.5 ± 1.0 50.1	30071.6 ± 355.6 29874.5
K27A hSB·hSB9	3.1 ± 0.2 3.5	50.2 ± 0.3 50.3	30333.7 ± 222.2 30412.6

Table 4.1: Structural Properties of wt and mutated complexes. Values of RMSD, of DCOM and of SASA as obtained by averaging the MD trajectories (left columns) and in the final MD snapshots (right columns). The averages are not taken for the *R28K* hSB·hSB9, *R201A* hSB·hSB9 and *R201K* hSB·hSB9, which are unstable during the dynamics.

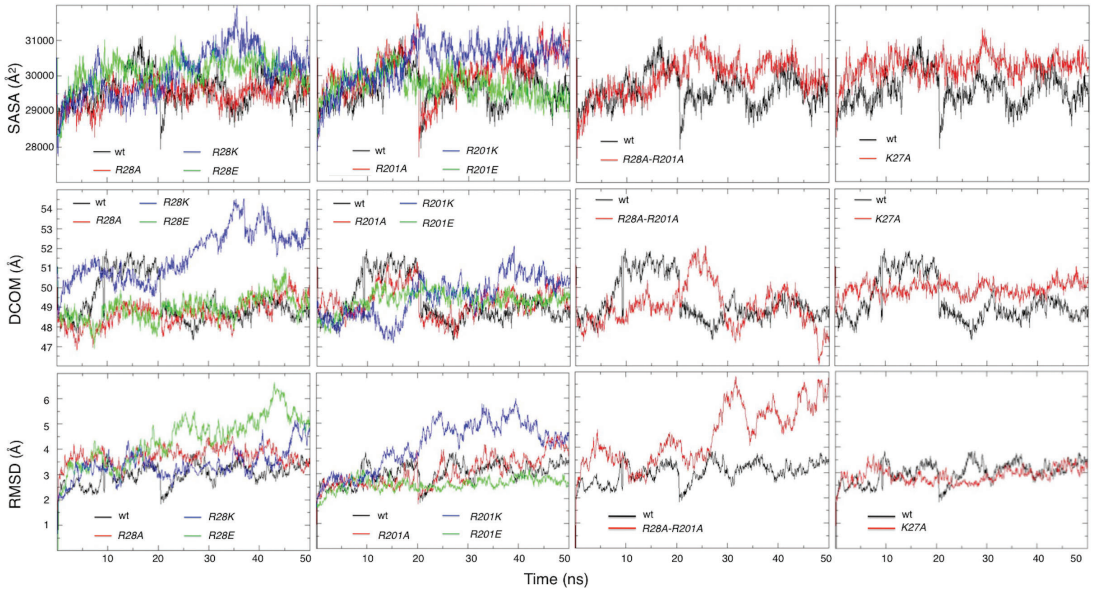

Figure 4.2: Structural properties. Structural properties. RMSD, DCOM and SASA (See text for a definition of these quantities) for wt and mutated hGB·hSB9 complexes, plotted as a function of simulated time.

For this purpose, we first used Baker's computational alanine scanning mutagenesis (CASM) procedure (28, 264), followed by molecular dynamics (MD) simulations.

Complex	Salt bridges	Hydrogen bonds	Hydrophobic contacts
wt hSB·hSB9	*K27*-E340	*Q23*-D258	*L26*-M343
	R28-E344	*N202*-S334	*P82*-A339
	K135-E266		*F85*-A339
	K180-E201		
R28A hSB·hSB9	*K27*-E340	*Y17*-M343	*L157*-V337
	K135-E266	*R201*-V337	*A28*-M343
	K180-E201	*N202*-S334	
R28E hSB·hSB9	*K83*-E340	*R201*-C335	*L157*-V337
	K180-E201	*N202*-E201	*F85*-V338
R201E hSB·hSB9	*K24*-D258	*R28*-M343	*P82*-A339
	K180-E201		*F85*-V338
	K135-E266		*L26*-M260
R28A-R201A hSB·hSB9	*K24*-D258	*S25*-S262	*A28*-M343
		L26-S262	*A201*-V337
		P132-S334	*F85*-V338
		N202-S334	
K27A hSB·hSB9	*K135*-E264	*G134*-E201	
	K135-E266		
	R28-E344		

Table 4.2: Contacts between hGB and hSB9. The Table reports salt bridges, hydrogen bonds and hydrophobic interactions in wt and *R28A*, *R28E*, *R201E*, *R28A-R201A* and *K27A* hGB·hSB9 complexes.

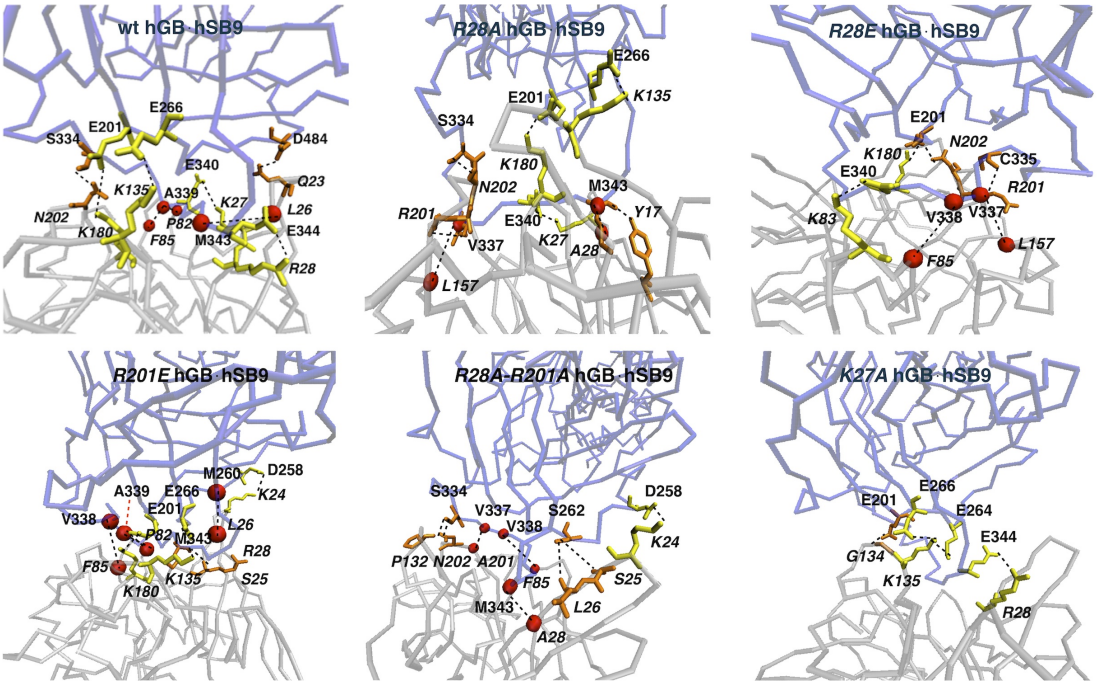

Figure 4.3: Protein-protein interfaces. Interfaces between hGB (grey) and hSB9 (blue) in the wt and mutated *R28A*, *R28E*, *R201E*, *R28A-R201A* and *K27A* hGB·hSB9 complexes which turned out to be stable in our MD simulations. Residues forming salt bridges (yellow sticks), hydrogen bonds (orange sticks) and hydrophobic interactions (red spheres) are highlighted.

CASM provides a rough estimate of the changes in the interaction free energy (DDG) between the two proteins upon mutation of hGB interface residues to Ala. The largest DDG values turned out to be associated with the *K27A* hGB·hSB9, *R28A* hGB·hSB9 and *R201A* hGB·hSB9 (Table A.1 in Appendix). Next, we performed 50 ns MD simulations of the mutations with the largest DDG along with others involving the same positions (*R28A-R201A* hGB·hSB9, *R28E* hGB·hSB9, *R28K* hGB·hSB9, *R201E* hGB·hSB9 and *R201K* hGB·hSB9) (Fig. A.3) in Appendix) in explicit solvent.

In ***R28A* hGB·hSB9**, the RMSD, DCOM and SASA are slightly larger than those of wt hGB·SB9 (Table 4.1, Fig. 4.2). The *K27*-E340, *K135*-E266 and *K180*-E201 salt bridges, present in wt hGB· hSB9, are formed as well in this case. The *R28A*-E344 salt bridge is lost (Table 4.2). The *N202*-S334 H-bond is maintained, while the *Q23*-D258 H-bond is lost. *Y17*, which does not form H-bonds in wt, in this case H-bonds

to M345. The *L26*-M343, *P82*-A339 and *F85*-A339 hydrophobic contacts, present in the wt, are absent here. However, M343 in this case forms a hydrophobic contact with *A28*. A new contact *L157*-V337 is formed as well (Table 4.2, Fig. 4.3). This complex therefore appears to be relatively stable during the time-scale simulated, although some of the interactions present in wt hGB·hSB9 are lost and some other absent in the wt hGB·hSB9 are formed.

In ***R28E* hGB·hSB9**, the RMSD, DCOM and SASA are larger than those of wt hGB·hSB9 (Table 4.1, Fig. 4.2). The *K135*-E266, *K27*-E340 and *R28*-E344 salt bridges are lost but E340 forms a new salt bridge with *K83*. *N202* form H-bonds to E201 instead of H-bonding to S334 as in wt hGB·hSB9. *R201* H-bonds to C335. The *Q23*-D258 salt bridge is also lost. *F85* forms a hydrophobic contact with V338 instead of A339, as in wt hGB·hSB9. *L157* forms a hydrophobic contact with V337 which is not present in wt hGB·hSB9 (Table 4.2). *L26* does not interact with hSB9's interface (Table 4.2) unlike in the wt hGB·hSB9.

We conclude that in spite of the increased DCOM and SASA values, the formation of new interactions stabilizes a rearranged hGB·hSB9 interface (Fig. 4.3). The complex is definitively stable in the timescale investigated.

In ***R28K* hGB·hSB9**, the RMSD, DCOM and SASA increase monotonically during the dynamics, pointing to a strong destabilization of the complex (Table 4.1, Fig. 4.2). Several stabilizing interactions at the interface are already lost in the time-scale investigated (data not shown). This complex is definitively unstable.

In ***R201A* hGB·hSB9**, the RMSD, COM's hGB·hSB9 distance and SASA increase with simulated time (Table 4.1, Fig. 4.2). Also in this variant, several interface interactions are lost at the end of the dynamics (data not shown). This complex is clearly unstable.

In ***R201E* hGB·hSB9**, RMSD, DCOM and SASA are comparable to the ones of the wt hGB·hSB9 complex (Table 4.1, Fig. 4.2). The *K27*-E340 and *R28*-E344 salt bridges are lost, but the *K180*-E201 and *K135*-E266 bridges are preserved and *K24*-D258 is newly formed (Table 4.2). *R28* forms a new H-bond with M343. This compensates for the loss of the *N202*-S334 bridge. The hydrophobic contacts are the same as in wt hGB·hSB9 (Table 4.2). The complex remains stable (Fig. 4.3).

In ***R201K* hGB·hSB9**, the RMSD, COM's hGB-SB9 distances and SASA increase monotonically during the entire dynamics (Table 4.1, Fig. 4.2). Several stabilizing

interactions at the interface are lost at the end of the MD (data not shown). In this case, the complex is unstable.

In ***R28A-R201A* hGB·hSB9**, the RMSD is 5.7 ± 0.6 Å. This relatively large value is caused by a series of conformational rearrangements (Fig. 4.3). However, these rearrangements are associated with the formation of a significant number of protein-protein contacts. S262 H-bonds to *S25* and *L26*. S334 H-bonds to the side chain of *N202* and the carbonyl of *P132* P132 backbone. *A28* and *A201* form hydrophobic contacts with M343 and V337, respectively. *F85* interacts with V338. Although protein/protein contacts involving *K27*, *R28A*, *K135* and *K180* are lost, *K24* forms an alternative salt bridge with D258. Consistently, DCOM and SASA do not vary largely (Table 4.1, Fig. 4.2). We conclude that this complex is rather stable. Indeed, the significant rearrangements observed are associated with new interactions at the interface (Table 4.2).

n ***K27A* hGB·hSB9**, the RMSD is similar to that of wt hGB·hSB9. The DCOM distance and the SASA fluctuate around a larger value than that of wt hGB·hSB9 (Table 4.1, Fig. 4.2). The salt bridges involving residue 27 and *K180* are obviously lost, while the ones at *R28* and *K135* are retained. However, a new hydrogen bond (between *G134* and E201) is formed. No hydrophobic interactions are present, in contrast to wt hGB·hSB9 (Table 4.2). However, despite the loss of interface contacts, the complex remains stable (Table 4.1, Fig. 4.3). Indeed, DCOM and SASA values do not significantly increase with time (Fig. 4.2).

Comparison with previous studies. The interface between hGB and full-length hSB9 in our model shows significant differences from an earlier structural prediction of hGB in complex with 12 residues (334-345) belonging to hSB9's reactive center loop. The latter is a solvent exposed stretch of amino acids representing the primary site of interaction with hGB (338). The differences are as follows: (i) in our complex, E201, D258 and E266 interact with residues of SB9 other than those of the loop modeled by (338); and (ii) our model features the *K27*-E340 and *R28*-E344 salt bridge, while in (338), the model features the *K27*-E344 salt bridge. *R28* and *K180* form salt bridges in our model (Table 4.2), while they form H-bonds in Sun et al's model (338). This suggests that inclusion of the entire hSB9 may be important to describe the interface.

Assays on enzymatic activity mutants in vitro. Taken together, the *in silico* results point to a few mutants featuring the largest destabilization in terms of loss of contacts as well as an increase in DCOM distance, SASA and RMSD: *R201A* hGB·hSB9, *R201K* hGB·hSB9 and *R28K* hGB·hSB9. In these mutants, the enzymatic activity of hGB (63) may not be too dissimilar from wt hGB activity because the active site is likely to exhibit the same structure of wt hGB. The relatively low predicted affinity of SB9 for hGB suggests that the activity of mutated hGB·hSB9 complexes will still be significant.

The other mutant complexes exhibit similar interacting surface (see DCOM and SASA values in Table 4.1 as well as a similar network of specific contacts at the interface (Table 4.2). In this case, enzymatic activity is predicted to be affected by hSB9 binding in a manner not too dissimilar from the behaviour of wt hGB·hSB9 (Fig. A.3 in Appendix).

However, it should be kept in mind that no quantitative activity predictions of enzymatic activity can be derived from classical molecular dynamics simulations and structural observations.

All proposed mutations were introduced into a hGB fusion protein by site-directed mutagenesis such that only single amino acids were exchanged. During the expression it was shown that all variants could be stably expressed in HEK293T cells and purified yielding the same amount of protein as wt hGB (data not shown).

We next measured the activity of wt and mutated hGB in absence and presence of recombinant hSB9. Enzymatic activity was detected by cleavage of the peptide Ac-IETD-*p*pNA, which mimics the cleavage site of the hGB substrate Pro-caspase 3 (63). We find that *R201A* hGB·hSB9, *R201K* hGB·hSB9 and *R28K* hGB·hSB9 in absence of hSB9 feature catalytic activity similar to that of WT hGB (Fig. 4.4). Enzymatic activity is significant also in the presence of hSB9, ranging from 46% *R201A* to 94% *R201K*. Our predictions are consistent with these results. The catalytic activity of the other mutated hGB·hSB9 complexes decreases relative to wt hGB, and is even lower in the case of the double mutant. Their remaining enzymatic activity after complexing ranges from 0.5% for the double mutant *R28A-R201A* to 25% for *R28E*. An exception is *R28A*, with a remaining activity of 54% in contrast to computational predictions (Fig. 4.4).

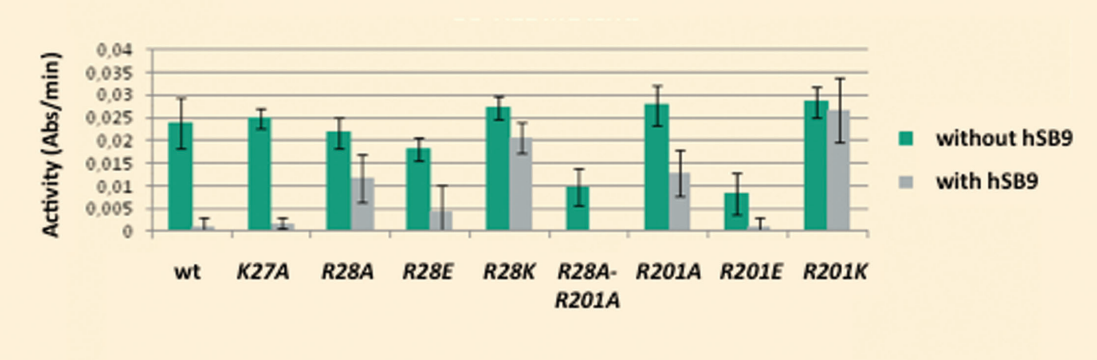

Figure 4.4: In vitro assays. Activity of hGB wt and mutant isoforms in presence (green histograms) or absence (grey histograms) of hSB9. The activity of the *K27A* hGB mutation without hSB9 was already measured by Sun et al (72).

In conclusion, only *R28K* hGB, *R201A* hGB, *R201K* hGB keep their enzymatic activity even in the presence of hSB9. Our predictions are fairly consistent with these experimental findings.

4.4 Concluding remarks.

We have presented a combined *in silico* and *in vitro* study aimed at designing mutants of hGB which retain their catalytic activity even in the presence of hSB9. The experimental data correlate fairly well with the computational predictions. In particular, *R28K* hGB, *R201A* hGB, *R201K* hGB did not lose their enzymatic activity and still retain most of it after incubation with the inhibitor hSB9. The mutant of choice for future therapeutic application is *R201K* hGB because, in the presence of hSB9, it preserves a comparable activity with wt hGB in the absence of the inhibitor (Fig. 4.4). Further *in vivo* studies with this candidate are required to test its activity in the cell, as well as potential immunogenicity and unspecific toxicity.

5

Investigation of PARP10-GSK3β interactions at the molecular level

5.1 Introduction

Carcinogenesis is a multistep process involving the progressive accumulation of genetic and epigenetic alterations that ultimately transform normal cells into neoplastic cells. The hallmarks of most of the cancers have been set in the discovery of mutations that produce oncogenes with dominant gain of function, such as Ras and c-MYC, involved in mitogenic cell signaling leading to increased cellular proliferation, and tumor suppressor genes with recessive loss of function, such as DNA repair factors and p53, mainly involved in the maintenance of genomic stability and cell cycle regulation (339, 340).

In recent years, several promising drug targets have been identified and novel drugs synthesized that target specific DNA repair proteins (341). These agents have shown impressive anti-cancer effects in preclinical studies in combination with classical cytotoxic treatments such as chemo- or radiotherapy.

Unfortunately, the efficacy of these is often impaired from the DNA repair mechanisms themselves (342, 343). One of the most advanced classes of DNA repair inhibitors is represented by inhibitors of poly-ADP-ribosylation (340). Poly-ADP-ribosylation is a

post-translational modification that occurs immediately after exposure of cells to DNA damaging agents. Poly-ADP-ribose polymerases (PARPs) catalyze the formation of a complex branched ADP-ribose polymer using NAD^+ as a donor of ADP-ribose residues (130). PARPs can covalently modify themselves (automodication) or other proteins (heteromodication). In humans, 17 genes and proteins have been identified, which are characterized by a conserved PARP signature motif. This motif is represented by the highly conserved amino acid sequence "HYE" that forms the active site (344). However, some of them seem to be catalytically inactive or possess mono-ADP-ribosylation rather than poly-ADP-ribosylation activity (87, 345). Thus, the PARP family is also referred to as diphtheria toxin-like ADP-ribosyltransferases (ARTDs) family (97, 346).

PARP10 is a peculiar member of the family in that it features only mono-ADP-ribosylation activity (34). It is localized in the nucleus and the cytoplasm. PARP10 was shown to act as a mono-ADP-ribosyltransferase that automodifies glutamate 882. Intriguingly, the glutamate of catalytic "HYE" triad, conserved across all the other PARP enzymes, is here missing. It is replaced by the isoleucine 987 (34) (Fig. 5.1a)

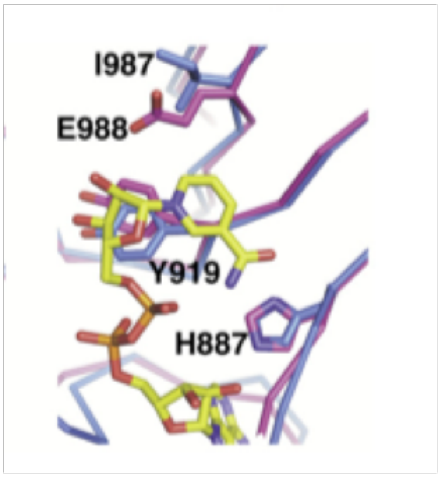

Figure 5.1: Superposition of PARP1 and PARP10 catalytic sites. Superposition of PARP1 and PARP10 catalytic sites, as shown by the X-ray structrures of the two proteins (347) (Karlberg et al., paper unpublished). In PARP10, the Glu988 is missing and is structurally replaced by the Ile987. From (34).

Hence, the catalytic mechanisms of PARP10 is different from those of the other PARP enzymes. The acidic target residue of the protein substrate has been proposed to substitute functionally the catalytic glutamate. This could be achieved by using substrate-assisted catalysis to transfer ADP-ribose (34).

In order to understand how PARP10 is regulated and in which physiological processes it is involved, it is crucial to point out which molecules PARP10 can modify and where the modification site is located.

Until now, PARP10 is only known to functionally associate with core histones and c-MYC (99). Thus, a protein microarray analysis was carried out to characterize new PARP10 interactors. 78 novel potential PARP10 substrates were identified with different functions (Fig. 5.2). 11 of them were validated with independent experiments and using different protein sources (Table 5.1).

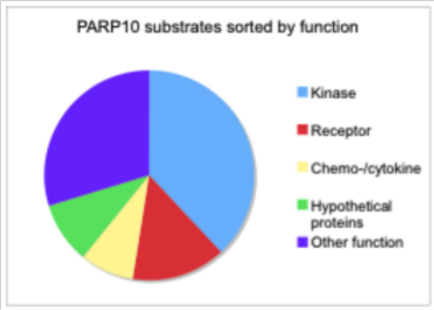

Figure 5.2: Identified PARP10 substrates.

Protein	Function
PDGFB	Growth factor
GSK3β	Kinase
IKKϵ	Kinase
KCNAB1	Voltage-gated channel
ACVR1	Serine/threonine kinase receptor
FES	Kinase
SRPK2	Kinase
CDK5	Kinase
p-TEFb	Kinase
DYRK1	Kinase
YY1	Transcription factor

Table 5.1: Validated PARP10 substrates. Validated PARP10 substrates identified with protein microarray experiments (internal communication, unpublished).

Pharmaceutically relevant is the fact that PARP-dependent ADP-ribosylation inhibits GSK3β serine-threonine kinase activity (Table 5.1) (109).[1] Indeed, the therapeutic potential of GSK3β inhibitors has become an important area of investigation in cancer chemotherapy (348). Moreover, GSK3β-mediated signaling has been shown to be involved in neuronal development (349), bipolar disorder (350), Alzheimer's (351) and Parkinson's disease (352).

Here, I have used computational methods to investigate GSK3β-PARP10 interactions at the molecular level, so far lacking. To guide the molecular docking, we have made the plausible (but as yet not demonstrated) assumption that one of GSK3β's Glu or Asp may functionally replace the missing equivalent in PARP10 of Glu988 of PARP1. Because experimental evidence (35, 36, 353) show that also positively charged residues from cellular partners play a crucial role in the catalysis, GSK3β's exposed Arg and Lys residues were also considered and included in the computational setup (see next section).
Previous preliminary experiments in Prof. Lüscher's lab ruled out Glu53, Glu121 and Glu279 from the pool of 76 possible candidates (354). On the basis of our calculations, additional 6 residues (Asp77, Glu211, Asp264, Glu290, Arg96, Lys197) were selected. Experiments carried out in Prof. Lüscher's lab were then performed to validate or disprove my predictions.

5.2 Computational methods.

5.2.1 Available structural information for the isolated moieties.

The GSK3β structure alone or in complex with diverse inhibitors have been solved by X-ray crystallography (355, 356). Here we consider the structure of the protein alone taken from PDB ID 1H8F (354).
The PARP10 double mutant structure in complex with the inhibitor 3-aminobenzamide (PDB ID 3HKV) (Karlberg et al., unpublished) is deposited in the Protein Data Bank

[1] The protein belongs to the glycogen synthase kinase subfamily. It is involved in many intracellular signaling systems. It functions in diverse cellular processes including proliferation, differentiation, motility and survival. Since its downstream targets include several key transcription factors such as c-MYC, it is regarded as an important modulator of cell physiology, including proliferation and death (33).

(PDB) (217). The PARP10 structure in complex with the NAD^+ cofactor has not yet been solved. However, recent computational studies (357) highlighted the presence of a common, scorpion-shaped structural motif recognizing the NAD^+ substrate. In particular, structural similarity has been found, despite the sequence diversity, between the NAD^+-binding motif of the bacterial diphtheria toxins and the structurally known ARTDs. 8 structures complexed with NAD^+ have been reported, with slightly different NAD^+ conformations (Fig. 5.3). The NAD^+ nicotinamide moiety shows conserved interactions with the scorpion motif and the adenine moiety a small conformational variability (357).

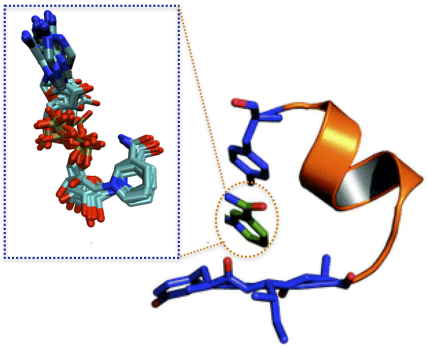

Figure 5.3: PARP10 scorpion motif. The nicotinamide moiety in conserved position is in the orange circle, the 8 NAD^+ conformations from (357) in the blue box.

5.2.2 Docking NAD^+ onto PARP10.

We predicted the structure of PARP10 in complex with NAD^+ by protein-ligand docking, based on the published structure of PARP10 (PDB ID 3HKV) in complex with the inhibitor 3-aminobenzamide (3AB). 3AB is an analog of the NAD^+ nicotinamide moiety and is located close to the catalytic residues H887 and Y919. In order to guide the docking procedure, we introduced two set of information: i) the interactions (hydrogen bonds and hydrophobic interactions) formed by 3AB in the PARP10 pocket, introduced as unambiguous distance restraints; ii) the common structural features shared by other poly-ADP-ribose polymerases and ADP-ribosylating toxins (357), described in the previous section. In particular, PARP10 catalytic residues His887 and Tyr919 and the NAD^+ were defined as active residues (see definition of active and passive residues in

Chapter 3 of this thesis).

The protein-ligand docking was performed using the software HADDOCK (for a description of the method, see Chapter 3 of this thesis) (21). The best structure belonging to the best cluster, in terms of HADDOCK scoring function (255) was retained. The HADDOCK scoring function takes into account several parameters, i.e. the van der Waals, electrostatics, restraint violation, desolvation and symmetry restraint energies and the buried surface area (358).

5.2.3 Docking PARP10/NAD^+ onto GSK3β.

The protein-protein docking between PARP10 and GSK3β was also performed using the software HADDOCK (21).
We used the active site residues His887 and Tyr919 and the NAD^+ as 'active residues' (21) for PARP10 in all of the HADDOCK runs. Each run was characterized by a different set of active residues for GSK3β. In particular, we used: i) all possible modification sites (Glu, Asp, Arg, Lys) characterized by a solvent accessibility of more than 40%, a requisite for the active residue to be enough exposed on the protein surface (21); ii) all possible pairs formed by one negatively charged (Asp or Glu) and one positively charged (Lys or Arg) residue, characterized by a structural distance ≤ 6 Å. The latter option was explored in order to take into account the possibility that the GSK3β catalytic Asp or Glu could be assisted by Lys or Arg as active residues for GSK3β (35, 36). A similar procedure as above was carried out to identify the best HADDOCK adduct. That was indeed identified in terms of HADDOCK scoring function, after having retained for each run the best structure belonging to the best cluster as in (359, 360).

5.2.4 Experimental methods (Carried out by M.Sc. Karla Fejis in the lab of Prof. Lüscher at the Medical School of RWTH Aachen University).

Site-directed mutagenesis. Primers were designed using Agilent Technologies' - QuickChange Primer Design tool and were ordered from Sigma. Phusion Polymerase (New England Biolabs) was used for the PCR according to manufacturers instructions with exception of the elongation time, which was increased to 1 minute per kb. 2 U Dpn1 (Fermentas) was added to the PCR reaction mixture afterwards and incubated

for at least one hour at 37° C, to digest all methylated, i.e. non-mutated DNA before heat shock transformation. All constructs created thus were sequenced to ensure successful mutagenesis without acquisition of any other mutations.

ADP-ribosylation assays. ADP-ribosylation assays were routinely carried out at 30C for 30 minutes unless indicated otherwise. The reaction mixture (50 mM Tris-HCl, pH 8.0, 0.2 mM DTT, 5 mM $MgCl_2$ and 50 μM β-NAD^+ (Sigma) and 1 μCi [^{32}P]-β-NAD^+ (Amersham Biosciences) was added to 1 μg enzyme in a total reaction volume of 30 μl. Reactions were stopped by adding SDS sample buffer and were subsequently boiled and run on SDS-PAGE. Incorporated radioactivity was analyzed by exposure of the dried gel to X-ray film. Samples used in subsequent kinase assays were cooled on ice before washing of beads with coupled GST-GSK3β in kinase assay buffer.

Kinase assays. Kinase buffer: 5 mM MOPS pH 7.2, 2.5 mM β-glycerophosphate, 1 mM EGTA, 0.4 mM EDTA, 4 mM $MgCl_2$, 50 μM DTT and 40 ng/μl BSA. [^{32}P]-γ-ATP was diluted to 0.16 μCi/mul in 250 μM ATP in 3x kinase assay buffer. 25ng GST-GSK3β or precipitated material was incubated in a reaction volume of 25 μl containing 5 μl 0.16 μCi/μl [^{32}P]-ATP solution and 5 μg substrate peptide RRRPASVPPSP-SLSRHS(pS)HQRR (Millipore), unless the substrate peptide was titrated in indicated concentrations. After incubating at 30° C for 15 minutes the reaction was stopped by placing on ice. Routinely 10 μl aliquots were spotted on P81 paper in duplicate, washed with 0.5% phosphoric acid and air-dried before scintillation counting. SDS-sample buffer was added to the samples analyzed by SDS-PAGE and subsequent autoradiography. Statistical significance was determined by employing two-sided student t-tests.

5.3 Results.

5.3.1 Structural prediction PARP10-NAD^+.

The best scoring structure derived from the ligand-protein procedure is represented in Fig. 5.4.

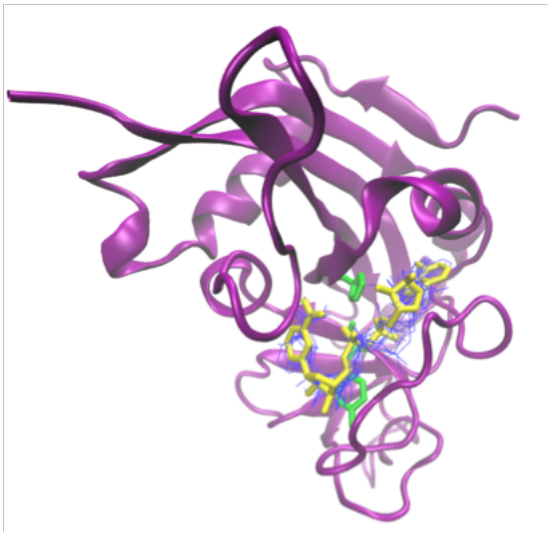

Figure 5.4: PARP10-NAD$^+$ complex. PARP10 (purple cartoon) in complex with NAD$^+$ as resulting from the docking procedure. The PARP10 catalytic residues His887 and Tyr919 are shown in green sticks, the known NAD$^+$ conformations (357) as blue lines.

In the docked NAD$^+$ pose, the nicotinamide moiety preserves the main interactions with the protein if compared (Fig. 5.4) with the other known NAD$^+$ included in the X-ray structures of the other proteins characterized by the same scorpion motif as PARP10 (357). These are: exotoxin A (361), diphteria toxin (362), ecto-ADP-ribosyltransferase 2.2 (363), cholera toxin (364), C3 exoenzyme (365), C3stau2 exoenzyme (366), VIP2 (367) and CDTa (368).

5.3.2 Docking PARP10-GSK3β.

Several plausible complex structures were predicted, in which a negatively charged residue and/or a positively charged residue were close to the PARP10 binding site.

According to the docking results, the following residues were chosen for preliminary experimental tests: Asp77, Lys197, Asp264 and Glu290, Arg96 and Glu211. Glu97 and Lys205 were also considered, due to the structural proximity with Arg96 and Glu211, respectively. The corresponding complexes arising from docking calculations are shown in Fig. 5.5.

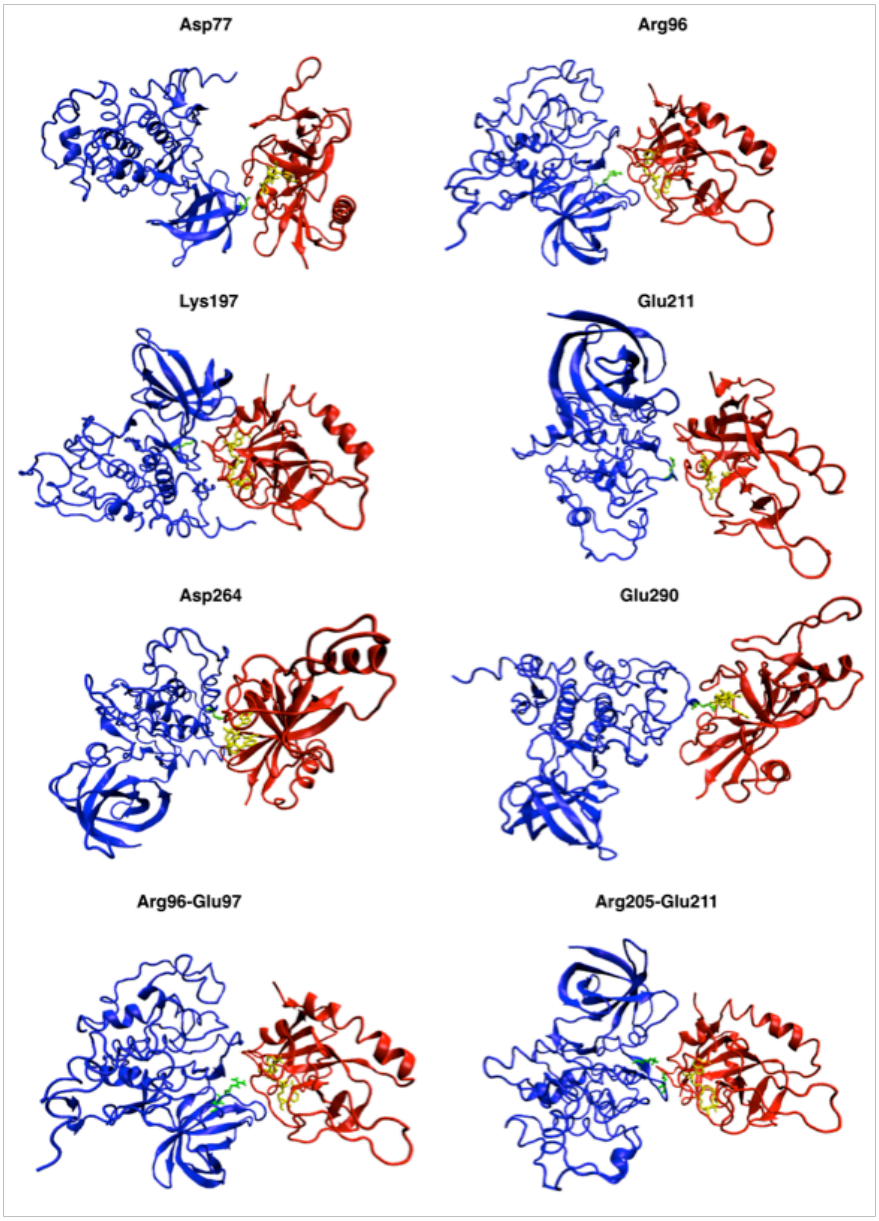

Figure 5.5: Docking poses. Docking poses for Asp77, Arg96, Lys197, Glu211, Asp264, Glu290, Arg96-Glu97 and Arg205-Glu211. GSK3β is represented in blue cartoons, PARP10 in red. NAD^+ is represented in yellow sticks, the potential modification site in green.

Limitation of the model. A limitation of this prediction is given by the fact that the mutations required to test the substrate could cause conformational changes. In this respect, large-scale molecular dynamics simulations could be performed. The

calculations may allow for structural rearrangements of the PARP10-GSK3β complex, leading to a more accurate interaction pose that observed with the docking procedure.

5.3.3 Mutations.

The Asp77, Arg96, Lys197, Glu211, Asp264, Glu290 single mutants and the Arg96-Glu97 and Arg205-Glu211 double mutants were expressed based on my predictions. If those GSK3β residues would play a role for the PARP10 catalytic activity, then these mutants should not be mono-ADP-ribosylated, thus retaining their kinase activity. Then, Karla Fejis in the lab of Prof. Lüscher carried out kinase and ADP-ribosylation *in vitro* assays on the GSK3β mutants (see Methods for details). It was found that they still get mono-ADP-ribosylated.

5.4 Discussion.

My biocomputing allowed for identifying 6 putative charged residues playing a role for PARP10/GSK3 interactions. None of them turned out to play a role for the binding in vitro. However, the implications for *in vivo* conditions should be taken with great care. Indeed:
i) different acceptor sites in different PARP enzymes have been reported in the literature (35, 36, 369). In particular:

- on PARP1 (35) and PARP2 (36) both glutamic acid and lysine residues are automodified;
- the recently identified mono-ADP-ribosyltransferase PARP15 was shown to modify neither acid nor basic residues, but threonine or serine instead (369);
- other mono-ADP-ribosyltransferases have been shown to modify also cysteines and arginines (353).

Although the substrate-assisted catalysis requires an acidic residue (aspartate or glutamate), these alternative mechanisms cannot be excluded in the case of PARP10 and thus, further investigations are required to clarify the nature of the GSK3β acceptor residue(s).

ii) glycation (370, 371) can modify lysines non-enzymatically, covering the modification site;
iii) experimental issues, such as the lack of antibodies against mono-ADP-ribosylation on specific residues and the lack of reliable mass spectrometry methods, as well as computational issues, such as the difficulty to obtain reliable docking results in blind calculations (i.e. without any experimental data related to the protein-protein interface) (18, 372) make the identification of modification sites a real challenge. In this context, a major drawback was represented by dealing with a protein (GSK3β) belonging to the extremely conserved (373) family of protein kinases. Further experiments would therefore greatly help identify a smaller interface region. This in turn would increase the reliability of docking predictions (21).

6

Molecular simulation of the NGF-TrkA complex.

6.1 Introduction

Neurotrophins are growth factors essential for the regulation of development, maintenance, and function of vertebrate nervous system (374) and for the modulation of synaptic transmission (128). The growth factor family comprises nerve growth factor (NGF), the brain-derived neurotrophic factor (BDNF), neurotrophin-3 (NT-3) and neurotrophin-4 (NT-4). NGF (121) is the founding member of the neurotrophins. Most of the signaling processes in the synaptic compartment are activated by NGF and its binding to the tropomyosin-related kinase A (TrkA) (37, 375) cognate receptor. NGF is involved in several disease states. The derangement of signaling processes involving NGF and TrkA is associated with the etiology of Alzheimer's disease (38, 120, 376, 377).

NGF/TrkA have been proposed to interact with a variety of cellular partners, including the low-affinity p75 receptor (378, 379) and the Abeta protein (376, 380). These interactions might play a role for the progress of Alzheimer's disease. Moreover, NGF-mediated TrkA activation typically leads to the regulation of synaptic strength and plasticity through a complex signaling pathway that involve several small G proteins, including Ras, Rap-1, and the Cdc-42-Rac-Rho family, as well as pathways regulated by MAP kinase, PI 3-kinase, phospholipase-C-gamma (PLC-), NFkB, and atypical protein kinase C pathways (124, 381, 382).

6. MOLECULAR SIMULATION OF THE NGF-TRKA COMPLEX.

Binding of proteins to their protein targets is expected to occur via two different molecular mechanisms (383, 384). In the "induced fit" model (383), binding of a protein to a ligand or another protein leads to a conformational change apt for optimal binding. In the "conformational selection" mechanism (384), some of the conformational states may have binding sites matching the shape and the chemistry of the interactor. The binding of the latter stabilizes these states with a "population shift" toward these conformers (385, 386). Mechanisms involving both a conformational selection step and an induced fit step have also been proposed (387, 388).

Here we use molecular simulations tools to investigate whether the conformational mechanism plays a role for NGF-TrkA recognition of cellular partners.
This complex is a dimer and features a "crab" shape with two extended pincers. Such "crab" body, of C_2-fold symmetry, is composed of the NGF homodimer flanked by the two pincers, each consisting of the five domains (D1-D5) of TrkA (40) (Fig. 6.1).

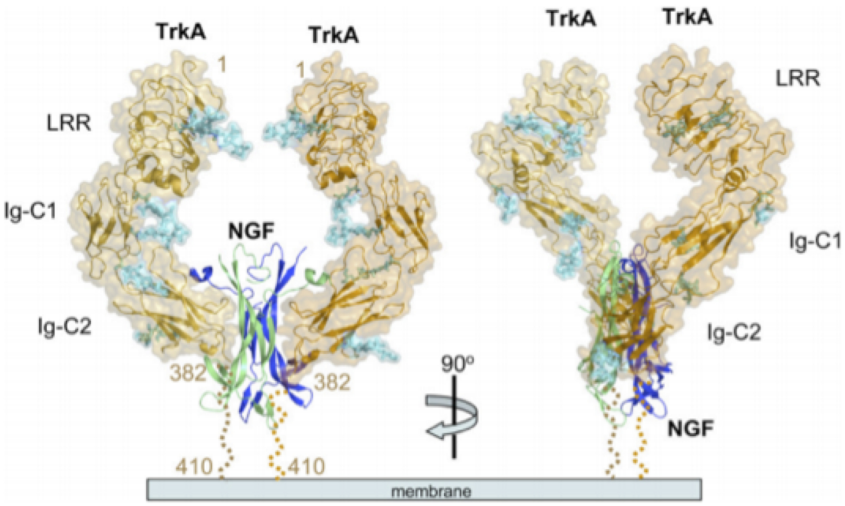

Figure 6.1: NGF structure complexed with TrkA extracellular domains (two side views). The TrkA extracellular domains are in orange and shown as ribbons with transparent molecular surface, and NGF is blue/green and only shown as ribbons. Cell membrane is depicted as a rectangular box. This picture has been taken from ref. (40).

We investigate structure, conformational fluctuations and energetics associated with the motion of the TrkA pincers in the structures by metadynamics (302) calculations. These calculations allow the accurate reconstruction of the free energy landscape of a biological system and the exploration of conformations separated by high free energy barriers.

6.2 Methods.

6.2.1 Equilibration.

The TrkA-NGF crystal structure was taken from ref. (40) (PDB ID 2IFG). The system underwent equilibration by classical MD, performed by using NAMD 2.7 (279). The overall charge of the system was neutralized by adding 14 Na^+ ions. The AMBER99SB (285) force field was used for the biomolecules and counter ions, and the TIP3P (256) force field was used for water molecules. Asparagine glycosylation sites (Asn95, Asn121, Asn188, Asn262, Asn281, Asn358) of Alpha-D-Mannose, Beta-D-Mannose, 2-(Acetylamino)-2-Deoxy-α-D-Glucopyranose and N-acetyl-D-Glucosamine were included in the simulations since deemed to stabilize the NGF/TrkA complex (40) and treated within the CHARMM36-all atom force field (389). Electrostatic interactions were calculated using the Particle Mesh Ewald method (390). The time-step was set to 2 fs. The SHAKE algorithm (391) was applied to fix all bond lengths.

6.2.2 Free energy calculations.

The free energy is calculated in terms of specific CVs, which describe the large-scale motions of TrkA's pincers. To define such CVs, we identify the NGF as the fulcrum of the pincer large scale motion. Next, we consider three regions of the pincers:
(i) the N-terminal regions (from Cys36 to Asn188, and from Cys36' to Asn188' (the symbol ' refers to the second pincer);
(ii) the central regions (from Ala189 to Glu339 and from Ala189' to Glu389');
(iii) the C-terminal regions (from Thr340 to Pro382 and from Thr340' to Pro382').

The CVs form therefore a fan-shaped set of descriptors (Fig. 6.2), defined by the three θ_1, θ_2, θ_3 bending angles:

$$\theta_i = a\cos\frac{(\mathbf{v}_{1i}\cdot\mathbf{v}_{2i})}{(|\mathbf{v}_{1i}||\mathbf{v}_{2i}|)} i = 1,3 \tag{6.1}$$

where $\mathbf{v}_{1i}$ and $\mathbf{v}_{2i}$ (i=1,3) are the three couples of vectors that connect the centers of each above mentioned TrkA regions with the NGF center of mass, as shown in Fig. 6.2.
θ_i depends on the orientation of the aforementioned three couple of vectors ($\mathbf{v}_{1i}$ and $\mathbf{v}_{2i}$ i=1,3) (Fig. 6.2). We consider as reference coordinates the X-ray structure of (40), where the values of θ_1, θ_2, θ_3 are 64°, 105° and 153°, respectively.

The free energy is calculated using 0.15 μs of well-tempered metadynamics within the framework of PLUMED (392). Gaussians of 0.2 kcal/mol height were deposited at one picosecond time interval, with a bias factor of 10. A temperature of 300 K was enforced using a Langevin thermostat. Four walkers were added in order to accelerate the free-energy convergence (393).
The free energy calculations were carried out by Dr. A. Pietropaolo (U. Catanzaro, Italy).

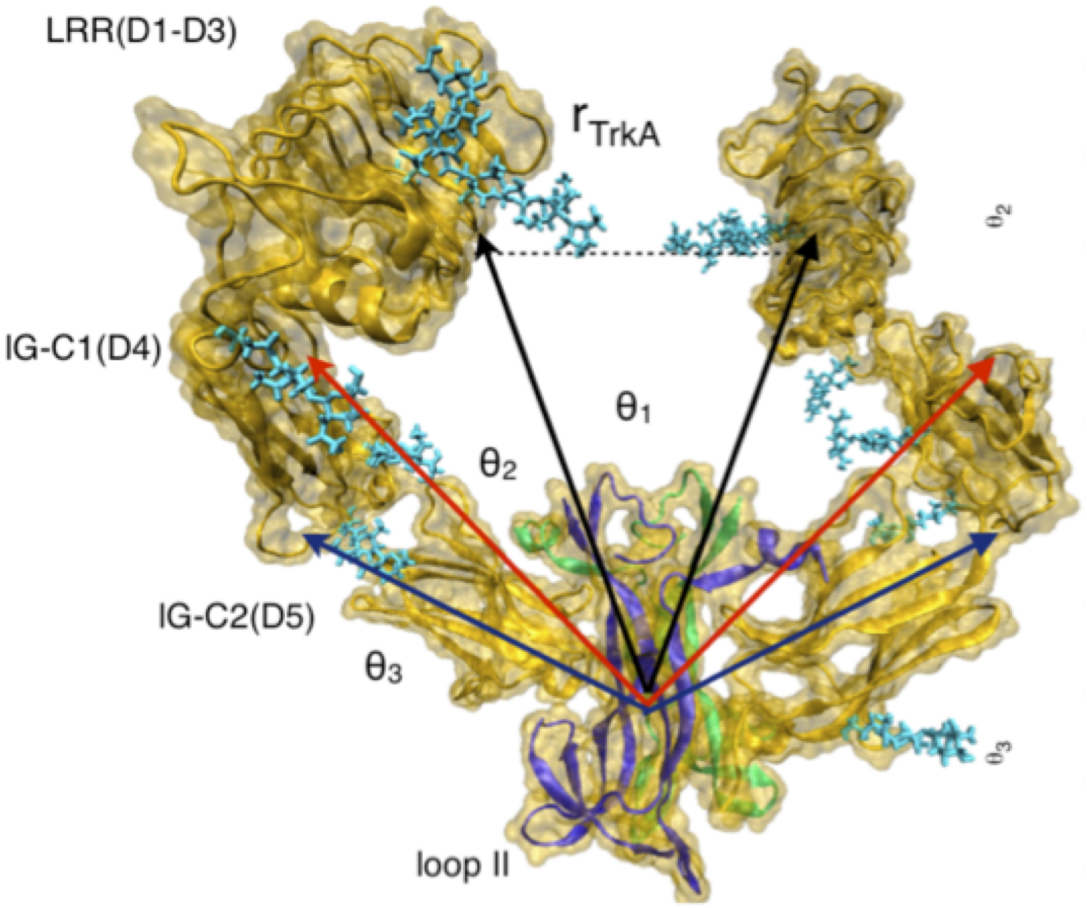

Figure 6.2: Collective variables (θ_1, θ_2, θ_3) for the pincer movement of NGF/TrkA complex (40). The collective variables are identified by the angles formed by the two pincers, highlighted by the black, blue and red lines, whose arrows point to the center of mass of each TrkA region.

6.3 Results

6.3.1 Structural features.

The calculations point to the presence of three, almost isoenergetic free energy valleys (separated by a barrier of 15 kcal/mol separates each valley) with close free energy basins (Fig. 6.3). One minimum (A in Fig. 6.3) correspond to the structure found through Xray diffraction by Wehrman et al. (40). It is centered at θ_1, $\theta_2 = 64°$, $105°$. The second minimum (B) features a root mean square deviation (RMSD) of 6 $\mathring{A}$ with respect to the previous one. It is centered at θ_1, $\theta_2 = 75°$, $110°$. It is characterized by a large extension of the two TrkA chains. The last two minima (C-D) feature values of θ_1, $\theta_2 = 80°$, $65°$ and $50°$, $60°$, respectively (Fig. 6.3). These conformations involve a large-scale rearrangment with respect to the X-ray structure (RMSD of 26 and 21 Å, respectively). As a result, NGF and TrkA are in closer proximity than in the other minima. These minima feature an asymmetric (C) and symmetric (D) displacement of the two TrkA chains (Fig. 6.3). The IgG-C1 subdomain is elongated with respect to A and B. This might corroborate the model by Zaccaro et al. (164), based on which the IgG-C1 subdomain is a hinge region regulating TrkA conformation and activation through large-scale conformational changes.

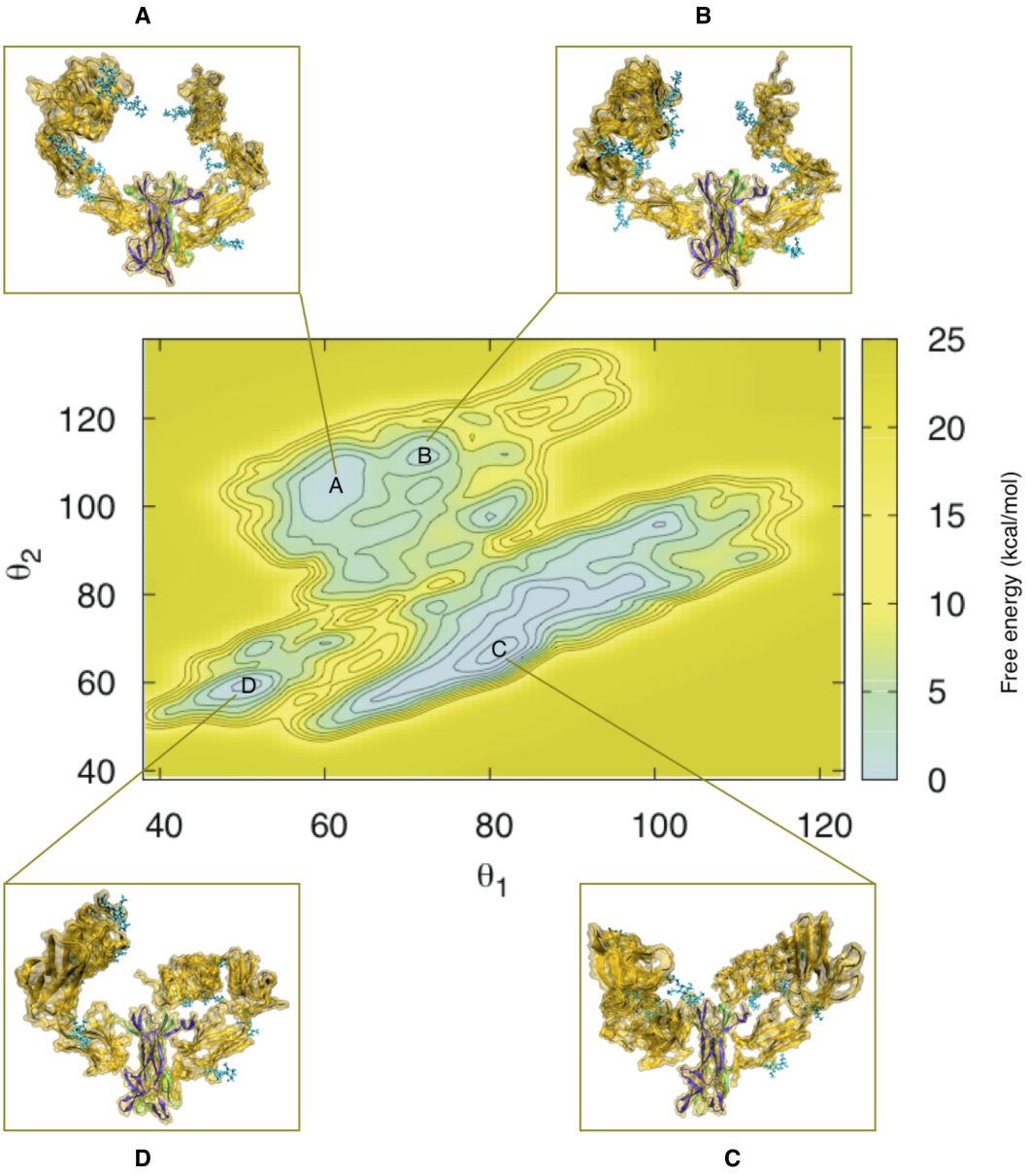

Figure 6.3: Free energy landscape of NGF/TrkA complex. Free energy basins are highlighted and shown by representative snapshots. NGF chains are shown in blue and green, TrkA chains are shown in gold and glycosilation sites are shown by light-blue sticks.

6.3.2 Validation with experimental data.

In order to validate our result with available experimental data, we calculated the theoretical CD spectrum (394) for the 4 minima using the parameters from (395) and compared it with the experimental spectrum (396).
The results point to a global preservation of the secondary structure with respect to the X-ray structure (40) also in farther minima C and D (Fig. 6.4a, 6.4b).

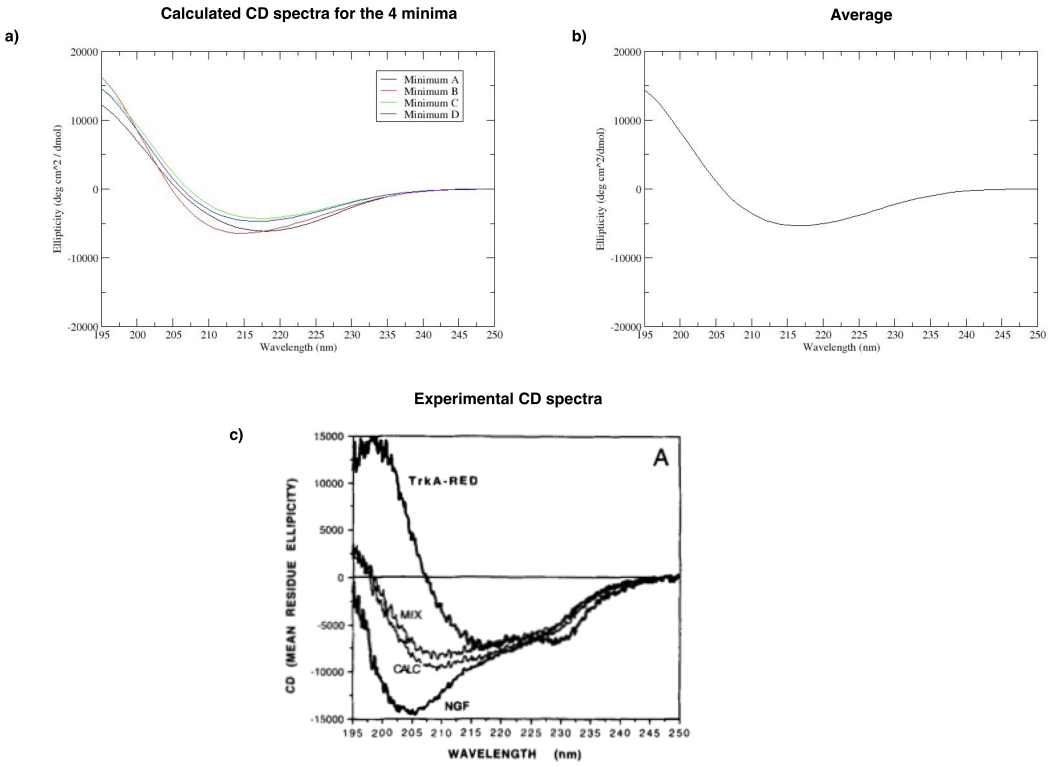

Figure 6.4: CD spectra. a) Calculated CD spectra for the four minima, and their average (b). c) Experimental NGF/TrkA CD spectrum (396). TrkA-RED indicates the experimental spectrum for TrkA alone, NGF the one for NGF alone, CALC is the average of the former two and MIX is the experimental spectrum for NGF/TrkA complex.

6.4 Discussion and conclusions.

Our calculations point out the existence of four isoenergetic minima of the NGF/TrkA complex. These minima eincompass as much as the 99.9 % of the total conformational space. Two of the minima (A, B in Fig. 6.3 turned out to be close to the initial,

X-ray structure (40). The other two (C, D in Fig. 6.3 feature instead significant conformational changes showing symmetric and asymmetric stretching of the Ig-C1 and LRR arms.

From one hand, this information might help structure-based drug design of peptide NGF mimics in conformations A-D, with possible applications in the fight against Alzheimer's and other neurodegenerative diseases (397, 398). Indeed, these mimic could act with either agonist or competitive antagonist on the TrkA receptor by exerting a neurotrophic effect (150).

From the other hand, our findings may have relevance for reported biological studies:
(i) Several experimental studies pinpointed the crucial role of NGF loop II (residues 40-49) for TrkA binding (163, 399, 400, 401). In the NGF/TrkA crystal structure (40), loop II does not interact with TrkA. However, the C-term of TrkA (residues 383-423) is missing in the X-ray structure (40, 163). In structures B, C, D, the orientation of the TrkA C-term with respect to NGF loop II changes significantly (Fig. 6.5). In particular, in B and D, one of the TrkA C-terms is located very close to the loop II of both NGF chains, suggesting that the interaction between TrkA and NGF loop II may be possible (Fig. 6.5b, 6.5d). This unprecedented finding provides a structural basis for the experimental observations in references (163, 399, 400, 401).

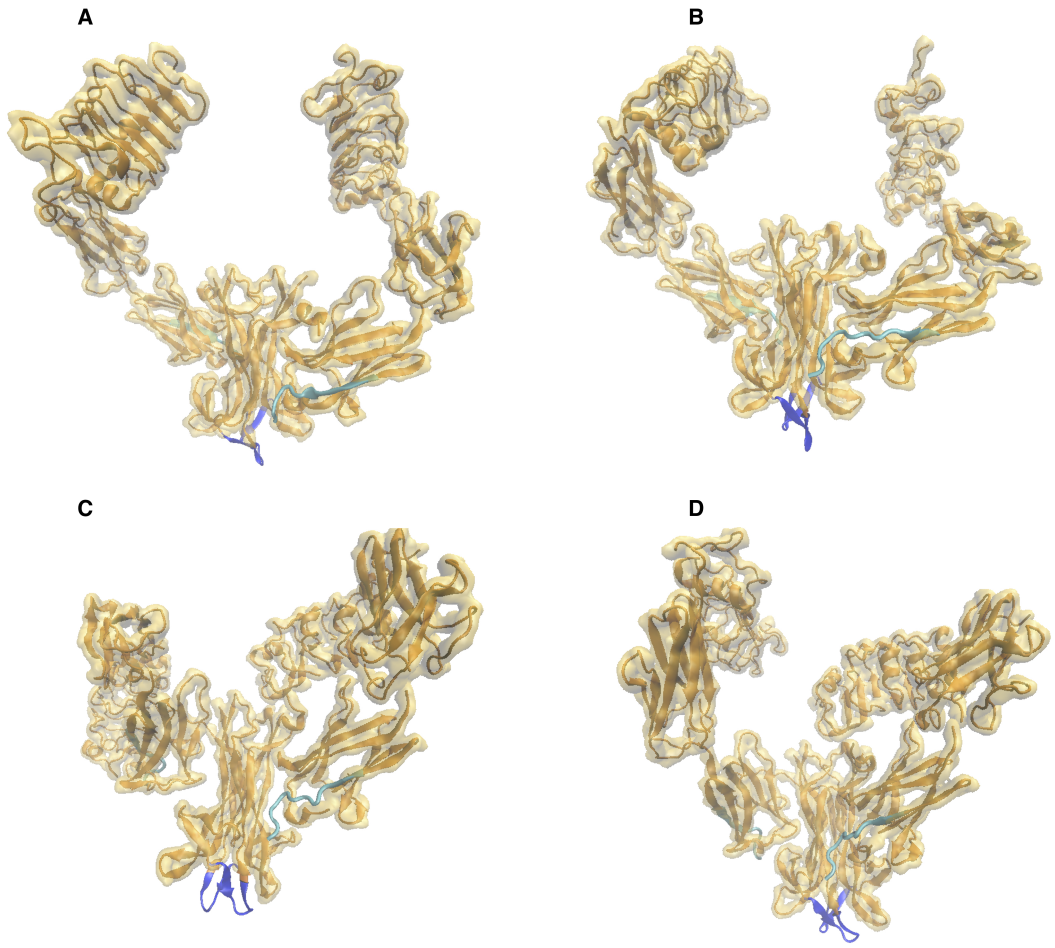

Figure 6.5: NGF-TrkA minima A, B, C, D (orange surface). The TrkA C-terms are displayed in cyan cartoons, the NGF loops II in blue cartoons.

(ii) *In vivo* and *in vitro* data indicate that p75 (a member of the tumor necrosis factor (TNF) receptor family (378) interacts with TrkA functionally (402, 403, 423), although likely depending on downstream signaling rather than physical complexes involving the two receptors (40). However, a very short-lived interaction between p75 and TrkA through the formation of a ternary complex might exist (379). At the speculative level, we suggest here that one or more than one minima A-D could play for the formation of this transient complex. NGF and TrkA have been shown to be involved in other 45 protein interactions (404). For some of them, the presence of different conformations A-D might play a role, albeit such role is speculative at present.

In conclusions our study shows that the NGF/TrkA complex adopts a different structure in water than in the crystal phase. This finding may have relevance for future ligand design studies and it does shed light on molecular biology studies on this pharmaceutically relevant system.

7

Conclusions

Providing molecular details of protein/protein interactions and integrating them in interaction networks may help understand greatly regulatory pathways, including those involved in human diseases (25). It also allows the design of protein-protein interaction inhibitors such as small molecules and peptides, which play an increasingly important role in pharmacology (405, 406).

Computational methods are an essential tool for structural proteomics studies. They include several techniques such as homology modeling, classical molecular dynamics, protein-protein docking and free energy calculations.

In this thesis, we investigated a variety of pharmacologically relevant protein complex systems. By using an arsenal of computational approaches we have faced challenges ranging from structural predictions to quantitative investigations of the energetics of complexes in aqueous solution.

I first focused on complexes involved in the development of optimized immunotoxins. These represent a novel therapeutic strategy against cancer (53). These immunotoxins are recombinantly built by attaching potent toxins (usually of bacterial origin) to antibody fragments that target cell-surface antigens (49). New generation constructs, based on human proteins, are able to prevent the formation of neutralizing antibodies (407).

In this context, an optimal strategy is based on human Granzyme B, a human serine protease with apoptotic activity able to kill tumor cell lines (31). Currently, the therapeutic usage of Granzyme B-based constructs is impaired by a major drawback, i.e. the *in vivo* protein inhibition by another serine protease, Serpin B9 (68). I rationally designed Granzyme B mutants able to show a reduced binding to Serpin B9 while retaining their catalytic activity. This was achieved by a structural prediction of the Granzyme B-Serpin B9 complex. This prediction was challenging because both the structure of Serpin B9 and the Granzyme B-Serpin B9 binding mode were not known. For this reason, we first predicted the Serpin B9 3D structure based on the analogous structure of another protein of the same family, Serpin B1. The next step was the structural prediction of the Granzyme B-Serpin B9 complex, based on the known structure of an analogous complex, the one between Serpin B9 and Trypsin. Finally, we identified the most important determinants of the interaction between the two proteins (the so-called hot spots) and observed the effect of their mutation along molecular dynamics simulations. Among the designed Granzyme B mutants, we selected three variants able to weaken the binding to Serpin B9. *In vitro* experiments performed by the group of Prof. Barth (Fraunhofer Institute, RWTH Aachen) validated our predictions. The three variants, in presence of Serpin B9, retained a catalytic activity comparable to the one of the isolated wild-type Granzyme B.

Then, we turned out attention to a system for which very limited experimental information is available. This is the complex between the GSK3β and PARP10 proteins involved in cancer pathways. PARP10 features mono-ADP-ribosylation activity (34). GSK3β may provide glutamate residue to perform the reaction (34). Here I carried out extensive and systematic protein-protein docking studies using this information as input in the calculations. The structural models of the complexes were tested against *in vitro* mutation and ADP-ribosylation assays performed by the group of Prof. Lüscher at the Clinics of the RWTH Aachen University. Ultimately this study, while leaving specific questions still open, showed the crucial importance of a fully integrated experimental and computational approach in unveiling the molecular details of protein-protein interactions.

We finally moved from a qualitative description to a quantitative energetics analysis of protein-protein interactions. We focused on the neurotrophin NGF, a growth factor essential for the nervous system function (374), and its receptor TrkA. The dysregulation of NGF-mediated signaling has been related to Alzheimer's disease (408), being this signaling required for the long-term survival of nucleus basalis cholinergic neurons (409). Understanding of the details of NGF/TrkA interaction represents a crucial step to obtain new insights into the downstream pathways linked to neurodegeneration. This latter work has been done in collaboration with Dr. Adriana Pietropaolo (U. Catanzaro, Italy). In particular, Dr. Pietropaolo carried out the free energy calculations presented here.

Free energy calculations on the NGF/TrkA X-ray complex (40) allowed to identify four isoenergetic conformers present in solution. Two of them are significantly different with respect to the known crystal structure (40). The calculations are in agreement with all the available experimental data. These include on one hand the circular dichroism experiments performed on the NGF/TrkA complex (396) and on the other hand all of the biochemical evidences (163, 399, 400, 401) which point to a functional role of NGF loop II binding to TrkA, which is not evident from the crystal structure (40). These results might help design rationally peptide mimics based on the previously unrecognized NGF/TrkA conformations. These mimics might function as NGF/TrkA signaling modulators, hence interfering with the derangement associated with the development of Alzheimer's disease.

Appendix A

Design of Human Granzyme B Variants resistant to Serpin B9.

A. DESIGN OF HUMAN GRANZYME B VARIANTS RESISTANT TO SERPIN B9.

Mutation	**ΔΔG (kcal/mol)**
Y17 hSB·hSB9	0.1
K24 hSB·hSB9	0.6
S25 hSB·hSB9	0.1
L26 hSB·hSB9	0.4
***K27* hSB·hSB9**	**1.2**
***R28* hSB·hSB9**	**1.5**
H44 hSB·hSB9	0.9
F85 hSB·hSB9	0.8
L133 hSB·hSB9	0.6
K135 hSB·hSB9	0.4
H136 hSB·hSB9	0.6
Y160 hSB·hSB9	0.4
Y199 hSB·hSB9	0.4
***R201* hSB·hSB9**	**1.6**

Table A.1: Alanine Scanning. Change in free energy calculations of the mutation to Ala of residues at the hGB·hSB9 interface in the complex, as obtained by using the Robetta webserver (28, 264). The top three mutations destabilizing the complex are in bold face. These values are very approximate and should be taken here only for relative comparisons.

Complex	**Water molecules**	Counterions
wt hSB·hSB9	27,615	7 Cl-
K27A hSB·hSB9	28,055	6 Cl-
R28A hSB·hSB9	28,055	6 Cl-
R28E hSB·hSB9	27,644	5 Cl-
R28K hSB·hSB9	27,995	7 Cl-
R28A-R201A hSB·hSB9	27,647	5 Cl-
R201A hSB·hSB9	28,055	6 Cl-
R201E hSB·hSB9	27,644	5 Cl-
R201K hSB·hSB9	27,615	7 Cl-

Table A.2: Details of the simulations. Number of water molecules and counterions added to the complexes.

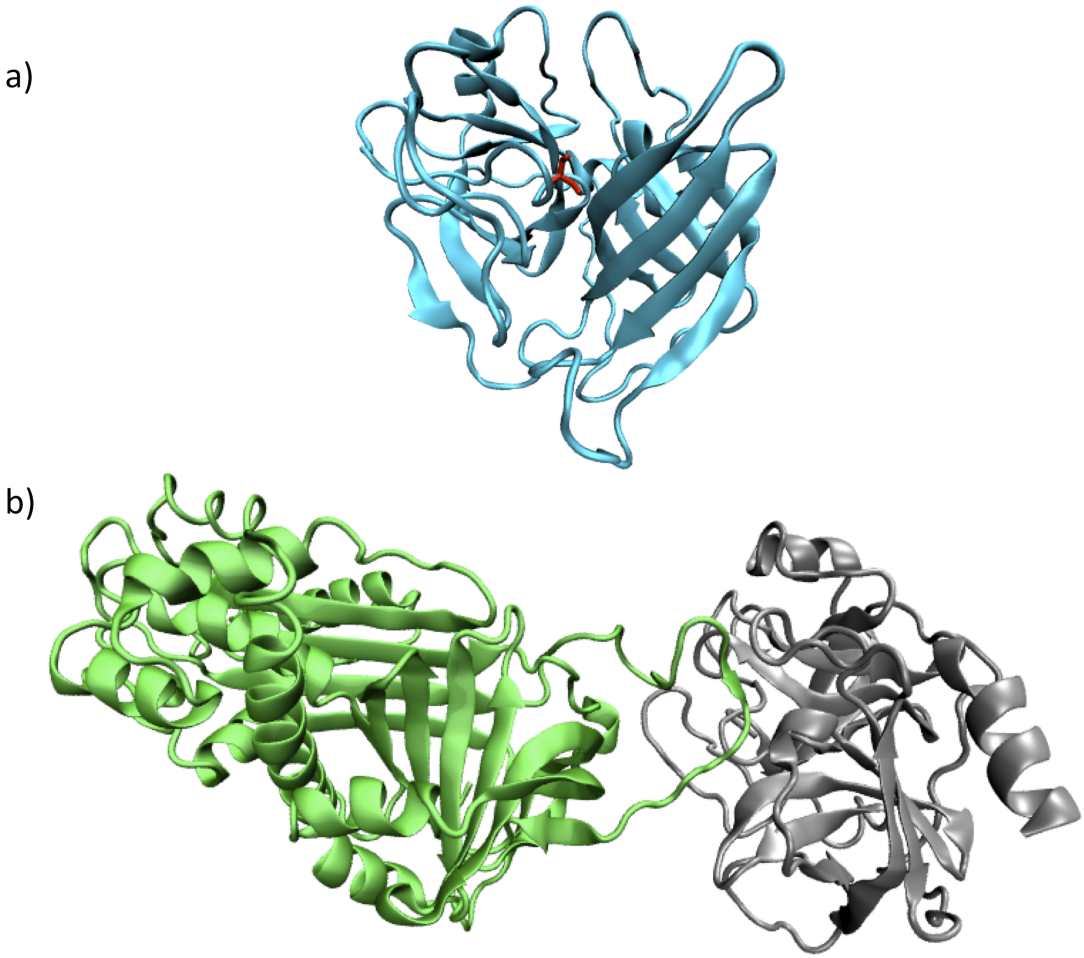

Figure A.1: Template for homology modeling. a) Cartoon of the hGB structure (67). The catalytic *S183* residue, which forms a covalent bond with P1 in S9 (321), is depicted in red sticks. (b) Cartoon of the rat Trypsin(silver)-*M.sexta* Serpin B1(green) complex structure (319) used as template for the structural prediction of the wt hGB·hSB9 complex.

A. DESIGN OF HUMAN GRANZYME B VARIANTS RESISTANT TO SERPIN B9.

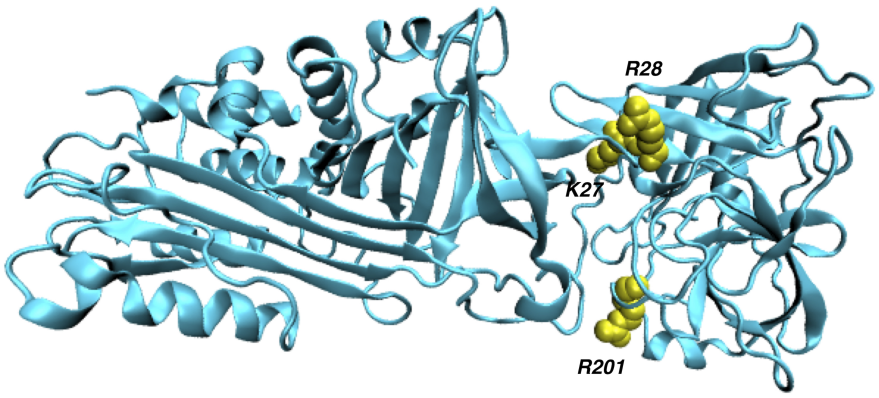

Figure A.2: Structure of wt hGB·hSB9 The hotspots in hGB (yellow spheres), emerging here from Baker's alanine scanning procedure (28, 264) using the Robetta webserver (Table A.1), are represented by yellow spheres.

Salt bridges	**Hydrogen bonds**	Hydrophobic contacts
D189-K353	S147-E191	F41-L355
	Q192-P352	F41-I356
	Q192-S354	L99-P352
	G193-K353	W215-I350
	A195-K353	W215-V351
	S214-K353	W215-P352
	Y39-L357	

Table A.3: Trypsin-Serpin B1 interactions. Contacts at the interface between rat Trypsin and *M. sexta* Serpin B1, as emerging from the X-ray structure (319).



Figure A.3: Structures of wt and mutated hGB·hSB9 complexes. a) wt hGB·hSB9 (thick line) and *R28K*, *R201A*, *R201K* hGB·hSB9; b) wt hGB·hSB9 and the other mutants (*R28A*, *R28E*, *R201E*, *R28A-R201A*, *K27A* hGB·hSB9). The last MD snapshots are shown for all systems.

A. DESIGN OF HUMAN GRANZYME B VARIANTS RESISTANT TO SERPIN B9.

Appendix B

NGF binding to p75 and downstream signaling pathways.

The three-dimensional structure of NGF in a complex with the extracellular domain of p75 (410) reveals an unexpected 2:1 NGF:p75 stoichiometry, with NGF appearing as a distorted dimer, which leads to the hypothesis that p75NTR binding to one side of NGF induces a bending of the entire NGF and disables the binding of a second p75NTR to the opposite site. This is in contrast with the complex between p75NTR and the neurotrophin NT3, where NT3 is a perfect homodimer, and the p75NTR receptors bind symmetrically to form a 2:2 cluster (411).

Historically, p75NTR was the first neurotrophin receptor to be identified as a NGF-binding transmembrane protein (412, 413). However, due to the lack of readily identifiable signal transducing modules, it was shown that p75NTR functions as a Trk co-receptor that increases its NGF binding affinity (414, 415).
As other members of the TNF receptors were demonstrated to modulate apoptotic responses (416), a potential role for p75NTR as a mediator of cell death was first hypothesized (417).
Indeed, *in vitro* analyses have shown that p75NTR induces cell death in hippocampal (418, 419) and sympathetic neurons (420, 421). In vivo, p75NTR plays a prominent role in apoptosis that occurs in neurons after traumatic injury (422) and has been implicated in developmental apoptosis in the peripheral nervous system (421). On the other hand, a large number of studies establish NGF-p75NTR interactions as a critical

biological event that governs life and death decisions in neuronal and non-neuronal cells (423). When p75NTR and TrkA are expressed in the same cell, however, survival signaling dominates (424).

In contrast, survival signaling predominantly reflects Trk activation (424). Upon NGF binding, adjacent TrkA receptors dimerize and undergo trans-autophosphorylation, followed by the activation of intracellular survival signaling cascades (128). The main events are represented by i) the activation of the phosphatidylinositol-3 (PI-3)-kinase/Akt pathway. These kinases were shown to be both necessary and sufficient for the survival of NGF-dependent sympathetic neurons (425); ii) the activation of the RAS/mitogen-activated protein (MAP) kinase cascade, leading to the phosphorylation of the transcription factor CREB and the activation of immediate-early genes (426).

References

[1] Srivastav R.K., Schwede S., Klaus M., Schwermann J., Gaestel M. and Niedenthal R. Monitoring protein-protein interactions in mammalian cells by trans-sumoylation. Biochem. J. 438: 495-503, 2011. 1

[2] Ryan D. and Matthews J. Protein-protein interactions in human disease. Curr. Opin. Struct. Biol. 2005, 15: 441-446. 1

[3] Fry D. C. Protein-protein interactions as targets for small molecule drug discovery. Biopolymers 2006, 84: 535-552. 1

[4] Uetz P., Giot L., Cagney G., Mansfield T.A., Judson R.S., Knight J.R., Lockshon D., Narayan V., Srinivasan M., Qureshi-Emili A., Li Y., Godwin B., Conover D., Kalbfleisch T., Vijayadamodar G., Yang M., Johnston M., Fields S. and Rothberg J.M. A comprehensive analysis of protein-protein interactions in Saccharomyces cerevisiae. Nature 403: 623-627, 2000. 1

[5] Ewing R.M., Chu P., Elisma F., Li H., Taylor P., Climie S., McBroom-Cerajewski L., Robinson M.D., OConnor L., Li M., Taylor R., Dharsee M., Ho Y., Heilbut A., Moore L., Zhang S., Ornatsky O., Bukhman Y.V., Ethier M. and Sheng Y. Large-scale mapping of human protein-protein interactions by mass spectrometry. Mol Syst Biol 3: 89, 2007. 1

[6] Ramani A.K., Li Z., Hart G.T., Carlson M.W., Boutz D.R. and Marcotte E.M. A map of human protein interactions derived from co-expression of human mRNAs and their orthologs. Mol Syst Biol 4: 180, 2008. 1

[7] Enright A.J,. Iliopoulos I., Kyrpides N.C. and Ouzounis C.A. Protein interaction maps for complete genomes based on gene fusion events. Nature 402: 86-90, 1999. 1

[8] Pellegrini M., Marcotte E.M., Thompson M.J., Eisenberg D. and Yeates T.O. Assigning protein functions by comparative genome analysis: protein phylogenetic profiles. Proc Natl Acad Sci USA 96: 4285-4288, 1999. 1

[9] Burger L. and van Nimwegen E. Accurate prediction of protein-protein interactions from sequence alignments using a Bayesian method. Mol Syst Biol 4: 165, 2008. 1

[10] Cusick M.E., Klitgord N., Vidal M. and Hill D.E. Interactome: gateway into systems biology. Hum. Mol. Genet. 14(2): 171-181, 2005. 1

[11] Karr J.R., Sanghvi J.C.m Macklin D.N., Gutschow M.V., Jacobs J.M., Bolival B., Assad-Garcia N., Glass J.I. and Covert M.W. A whole-cell computational model predicts phenotype from genotype. Cell 150(2): 389-401. 1

[12] Aloy P. and Russell R.B. Interrogating protein interaction networks through structural biology. PNAS 99(9): 5896-5901, 2002. 1

[13] Schmidt B. and Hogg P.J. Search for allosteric disulfide bonds in NMR structures. BMC Struct. Biol. 7: 49, 2007. 1

[14] Robinson C.V., Sali A. and Baumeister W. Molecular sociology of the cell. Nature 450: 973-982, 2007. 1

[15] Yee A.A., Savchenko A., Ignachenko A., Lukin J., Xu X., Skarina T., Evdokimova E., Liu C.S., Semesi A., Guido V., Edwards A.M. and Arrowsmith C.H. NMR and X-ray crystallography, complementary tools in structural proteomics of small proteins. J Am Chem Soc. 127(47): 16512-16517, 2005. 1

[16] Bonvin A.M.J.J. Flexible proteinprotein docking. Curr. Opin. Struct. Biol. 16: 194, 2006. 1

[17] Montalvao R.W., Cavalli A., Salvatella X., Blundell T.L. and Vendruscolo M. Structure Determination of Protein-Protein Complexes Using NMR Chemical Shifts: Case of an Endonuclease Colicin-Immunity Protein Complex. J. Am. Chem. Soc., 130 (47): 15990-15996, 2008. 2

[18] Lensink M.F. and Wodak S.J. Blind predictions of protein interfaces by docking calculations in CAPRI. Proteins 78: 3085-3095, 2010. 2, 82

[19] Wiehe K., Pierce B., Tong W.W., Hwang H., Mintseris J., et al. The performance of ZDOCK and ZRANK in rounds 6-11 of CAPRI. Proteins 69: 719-725, 2007. 2

[20] Comeau S.R., Gatchell D.W., Vajda S. and Camacho C.J. ClusPro: an automated docking and discrimination method for the prediction of protein complexes. Bioinformatics 20: 45-50, 2004. 2

[21] Dominguez C., Boelens R. and Bonvin A.M. HADDOCK: a protein-protein docking approach based on biochemical or biophysical information. J Am Chem Soc 125: 1731-1737, 2003. 2, 39, 41, 77, 82

[22] Wang C., Bradley P. and Baker D. Protein-protein docking with backbone flexibility. J Mol Biol 373: 503-519, 2007. 2, 38

[23] Morillas M., Gomez-Puertas P., Rubi B., Clotet J., Arino J., Valencia A., Hegardt F.G., Serra D. and Asins G. Structural model of a malonyl-CoA-binding site of carnitine octanoyltransferase and carnitine palmitoyltransferase I: mutational analysis of a malonyl-CoA affinity domain. J. Biol. Chem. 277: 11473-11480, 2002. 2

[24] Dobrodumov A. and Gronenborn A.M. Filtering and selection of structural models: combining docking and NMR. Proteins 53: 18-32, 2003. 2

[25] Tuncbag N., Gursoy A. and Keskin O. Prediction of proteinprotein interactions: unifying evolution and structure at protein interfaces. Phys. Biol. 8, 2011. 2, 95

REFERENCES

[26] Ma B., Elkayam T., Wolfson H. and Nussinov R. Protein-protein interactions: structurally conserved residues distinguish between binding sites and exposed protein surfaces. PNAS 100: 5772-5777, 2003. 2

[27] Yao H., Kristensen D.M., Mihalek I., Sowa M.E., Shaw C., Kimmel M., Kavraki L. and Lichtarge O. An accurate, sensitive and scalable method to identify functional sites in protein structures. J. Mol. Biol. 326: 255-261. 2003. 2

[28] Kortemme T. and Baker D. A simple physical model for binding energy hot spots in protein-protein complexes. Proc. Natl. Acad. Sci. USA 99(22): 14116-14121, 2002. 2, 42, 58, 61, 64, 100, 102

[29] Kamisetty H., Ramanathan A., Bailey-Kellogg C. and Langmead C.J. Accounting for conformational entropy in predicting binding free energies of protein-protein interactions. Proteins 79(2): 444-462, 2011. 2

[30] Meirovitch H. Recent developments in methodologies for calculating the entropy and free energy of biological systems by computer simulation. Curr Opin Struct Biol 17(2): 181-186, 2007. 2

[31] Rosenblum M.G. and Barth S. Development of novel, highly cytotoxic fusion constructs containing granzyme B: unique mechanisms and functions. Curr, Pharm. Des. 15(23): 2676-2692, 2009. 2, 57, 96

[130] Yung T.M., Sato S. and Satoh M.S. Poly(ADP-ribosyl)ation as a DNA damage-induced post-translational modification regulating poly(ADP-ribose) polymerase-1-topoisomerase I interaction. J Biol Chem. 279(38): 39686-39696, 2004. 3, 72

[33] Jope R.S. and Bijur G.N. Mood stabilizers, glycogen synthase kinase-3beta and cell survival. Mol Psychiatry Suppl 1: S35-45, 2002. 3, 75

[34] Kleine H., Poreba E., Lesniewicz K., Hassa P.O., Hottiger M.O., Litchfield D.W., Shilton B.H. and Lscher B. Substrate-assisted catalysis by PARP10 limits its activity to mono-ADP-ribosylation. Mol Cell. 32(1): 57-69, 2008. 3, 15, 16, 72, 73, 96

[35] Altmeyer M., Messner S., Hassa P.O., Fey M. and Hottiger M.O. Molecular mechanism of poly(ADP-ribosyl)ation by PARP1 and identification of lysine residues as ADP-ribose acceptor sites. Nucleic Acids Res. 37(11): 3723-3738, 2009. 3, 75, 77, 81

[36] Haenni S.S., Hassa P.O., Altmeyer M., Fey M., Imhof R. and Hottiger M.O. Identification of lysines 36 and 37 of PARP-2 as targets for acetylation and auto-ADP-ribosylation. Int J Biochem Cell Biol. 40(10): 2274-2283, 2008. 3, 75, 77, 81

[37] Kaplan D.R., Hempstead B.L., Martin-Zanca D., Chao M.V. and Parada L.F. The trk proto-oncogene product: a signal transducing receptor for nerve growth factor. Science 252: 554-558, 1991. 3, 83

[38] Matrone C., Barbagallo A.P.M., La Rosa L.R., Florenzano F., Ciotti M.T., APP is Phosphorylated by TrkA and Regulates NGF/TrkA Signaling. J. Neurosci. 31: 11756-11761, 2011. 4, 83

[39] Calissano P., Matrone C. and Amadoro G. Nerve growth factor as a paradigm of neurotrophins related to Alzheimer's disease. Dev Neurobiol. 70(5): 372-383, 2010. 4

[40] Wehrman T., He X., Raab B., Dukipatti A., Blau H. and Garcia K.C. Structural and Mechanistic Insights into Nerve Growth Factor Interactions with the TrkA and p75 Receptors. Neuron. 53(1): 25-38, 2007. viii, 4, 23, 24, 25, 84, 85, 86, 87, 88, 90, 91, 93, 97

[41] Jemal A., Bray F., Center M.M., Ferlay J., Ward E. and Forman D. Global cancer statistics. CA: A cancer journal for Clinicians. 61(2): 69-90, 2011. 7

[42] Anand P., Kunnumakkara A.B., Sundaram C., Harikumar K.B., Tharakan S.T., Lai O.S., Sung B. and Aggarwal B.B. Cancer is a preventable Disease that Requires Major Lifestyle Changes. Pharm. Res. 25(9): 2097-2116, 2008. 7

[43] Kinzler K.W. and Vogelstein B. Introduction. The genetic basis of human cancer(2nd, illustrated, revised ed.). New York: McGraw-Hill, Medical Pub. Division. p. 5. ISBN 978-0-07-137050-9, 2002. 7

[44] Gherardi E., Birchmeier W., Birchmeier C. and Woude G.V. Targeting MET in cancer: rationale and progress. Nature Reviews Cancer 12: 89-103, 2012. 7

[45] Jones K.L. and Buzdar A.U. A review of adjuvant hormonal therapy in breast cancer. Endocrine Related Cancer 11: 391-406, 2004. 7

[46] Hayden C.E. Cutting off cancers supply lines. Nature 458, 686-687, 2009. 7

[47] Mellman I., Coukos G. and Dranoff G. Cancer immunotherapy comes of age. Nature 480: 480-489, 2011. 7

[48] Zhukov N.V. and Tjulandin S.A. Targeted therapy in the treatment of solid tumors: practice contradicts theory. Biochemistry (Mosc) 73(5): 605-618, 2008. 7

[49] Pastan I., Hassan R., Fitzgerald D.J. and Kreitman R.J. Immunotoxin therapy of cancer. Nat Rev Cancer. 6(7): 559-565, 2006. 8, 95

[50] Teicher B.A. Antibody-drug conjugate targets. Curr Cancer Drug Targets. 9(8): 982-1004, 2009. 8

[51] Kreitman R.J. Recombinant immunotoxins for the treatment of chemoresistant hematologic malignancies. Curr. Pharm. Des. 15(23): 2652-2664, 2009. 8

[52] Hetzel C., Bachran C., Fischer R., Fuchs H., Barth S. and Stocker M. Small cleavable adapters enhance the specific cytotoxicity of a humanized immunotoxin directed against CD64-positive cells. J. Immunother. 31(4): 370-376, 2008. 8

[53] Pastan I., Hassan R., FitzGerald D.J. and Kreitman R.J. Immunotoxin treatment of cancer. Annu Rev Med. 58 :221-237, 2007. 8, 58, 95

[54] Vitetta E.S. Immunotoxins and vascular leak syndrome. Cancer J. Scient. Am. 6(Suppl.): S218-S224, 2000. 8

[55] Mathew M. and Verma R.S. Humanized immunotoxins: a new generation of immunotoxins for targeted cancer therapy. Cancer Sci. 100(8): 1359-1365, 2009. 9

[56] Kurschus F.C. and Jenne D.E. Delivery and therapeutic potential of human granzyme B. Immun. Rev 235(1): 159-171, 2010. 9, 58

[57] Grossman, W.J. and Revell P.A The orphan granzymes of humans and mice. Curr Opin Immunol 15(5): 544-552, 2003. 9, 58

[58] Raja S.M. and Wang B. Cytotoxic cell granule-mediated apoptosis. Characterization of the macromolecular complex of granzyme B with serglycin. J Biol Chem 277(51): 49523-49530, 2002. 9

[59] Trapani J.A. and Sutton V.R. Granzyme B: pro-apoptotic, antiviral and antitumor functions. Curr Opin Immunol 15(5): 533-543, 2003. 9

[60] Andrade F. and Roy S. Granzyme B directly and efficiently cleaves several downstream caspase substrates: implications for CTL-induced apoptosis. Immunity 8(4): 451-460, 1998. 9, 58

[61] Trapani J.A. and Jans D.A. Efficient nuclear targeting of granzyme B and the nuclear consequences of apoptosis induced by granzyme B and perforin are caspase-dependent, but cell death is caspase-independent. J. Biol. Chem. 273(43): 27934-27938, 1998. 9

[62] Barry M. and Heibein J.AGranzyme B short-circuits the need for caspase 8 activity during granule-mediated cytotoxic T-lymphocyte killing by directly cleaving Bid. Mol Cell Biol 20(11): 3781-3794, 2000. 9

[63] Metkar S.S. and Wang B. Granzyme B activates procaspase-3 which signals a mitochondrial amplification loop for maximal apoptosis. J Cell Biol 160(6): 875-885, 2003. 9, 62, 69

[64] Lieberman J. The ABCs of granule-mediated cytotoxicity: new weapons in the arsenal. Nature Rev Immunol 3: 361-370, 2003. 9, 10

[65] Porter A.G. and Janicke R.U. Emerging roles of caspase-3 in apoptosis. Cell Death Differ. 6 (2): 99-104, 1999. 9

[66] Estebanez-Perpina E., Fuentes-Prior P., Belorgey D., Braun M., Kiefersauer R., Maskos K., Huber R., Rubin H. and Bode W. Crystal structure of the caspase activator human granzyme B, a proteinase highly specific for an Asp-P1 residue. Biol Chem 381(12): 1203-1214, 2000 10, 60

[67] Rotonda J., Garcia-Calvo M., Bull H.G., Geissler W.M., McKeever B.M., Willoughby C.A., Thornberry N.A. and Becker J.W. The three-dimensional structure of human granzyme B compared to caspase-3, key mediators of cell death with cleavage specificity for aspartic acid in P1. Chem Biol 8(4): 357-368, 2001. 10, 58, 60, 101

[68] Bird C.H. and Sutton V.R. Selective regulation of apoptosis: the cytotoxic lymphocyte serpin proteinase inhibitor 9 protects against granzyme B-mediated apoptosis without perturbing the Fas cell death pathway. Mol Cell Biol 18(11): 6387-6398, 1998. 12, 58, 63, 96

[69] Bladergroen B.A. and Meijer et al. Expression of the granzyme B inhibitor, protease inhibitor 9, by tumor cells in patients with non-Hodgkin and Hodgkin lymphoma: a novel protective mechanism for tumor cells to circumvent the immune system? Blood 99(1): 232-237, 2002. 12, 58

[70] van Houdt I.S. and Oudejans J.J. Expression of the apoptosis inhibitor protease inhibitor 9 predicts clinical outcome in vaccinated patients with stage III and IV melanoma. Clin Cancer Res 11(17): 6400-6407, 2005. 12

[71] Bots M., Van Bostelen L., Rademaker M.T., Offringa R. and Medema J.P. Serpins prevent granzyme-induced death in a species-specific manner. Immunol Cell Biol 84(1): 79-86, 2006. 12

[72] Sun J., Bird C.H., Sutton V., McDonald L., Coughlin P.B., De Jong T.A., Trapani J.A. and Bird P.I. A cytosolic granzyme B inhibitor related to the viral apoptotic regulator cytokine response modifier A is present in cytotoxic lymphocytes. J Biol Chem 271(44): 27802-27809, 1996. 12, 58, 63, 70

[73] Bladergroen B.A., Strik M.C., Bovenschen N., van Berkum O., Scheffer G.L., Meijer C.J., Hack C.E. and Kummer J.A. The granzyme B inhibitor, protease inhibitor 9, is mainly expressed by dendritic cells and at immune-privileged sites. J Immunol 166(5): 3218-3225. 12, 58

[74] Hirst T.R. and Holmgren J. Conformation of protein secreted across bacterial outer membranes: a study of enterotoxin translocation from Vibrio cholerae. Proc. Natl Acad. Sci. USA 84, 7418-7422, 1987. 12

[75] ten Berge R.L., Meijer C.J., Dukers D.F., Kummer J.A., Bladergroen B.A., Vos W., Hack C.E., Ossenkoppele G.J. and Oudejans JJ. Expression levels of apoptosis-related proteins predict clinical outcome in anaplastic large cell lymphoma. Blood 99(12): 4540-4546, 2002. 12

[76] Rousalova I., Krepela E., Prochazka J., Cermak J. and Benkova K. Expression of proteinase inhibitor-9/serpinB9 in non-small cell lung carcinoma cells and tissues. Int J Oncol 36(1): 275-283, 2010. 12

[77] Ray M., Hostetter D.R., Loeb C.R., Simko J. and Craik C.S. Inhibition of Granzyme B by PI-9 protects prostate cancer cells from apoptosis. Prostate. doi: 10.1002/pros.21486, 2011. 12

[78] Liu Y. and Cheung L.H. Mechanistic studies of a novel human fusion toxin composed of vascular endothelial growth factor (VEGF)121 and the serine protease granzyme B: directed apoptotic events in vascular endothelial cells. Mol Cancer Ther 2(10): 949-959, 2003. 12

[79] Stahnke B. and Thepen T. Granzyme B-H22(scFv), a human immunotoxin targeting CD64 in acute myeloid leukemia of monocytic subtypes. Mol. Cancer Ther. 7(9): 2924-2932, 2008. 12, 62

[80] Zhang L. and Zhao J. HER2-targeting recombinant protein with truncated pseudomonas exotoxin A translocation domain efficiently kills breast cancer cells. Cancer Biol Ther 7(8): 1226-1231, 2008. 12

REFERENCES

[81] Kanatani I., Lin X., Yuan X., Manorek G., Shang X., Cheung J.H., Rosenblum M.G. and Howell S.B. Targeting granzyme B to tumor cells using a yoked human chorionic gonadotropin. Cancer Chemother Pharmacol 68(4): 979-990, 2011. 12

[82] Irving J.A., Pike R.N., Lesk A.M. and Whisstock J.C. Phylogeny of the serpin superfamily: implications of patterns of amino acid conservation for structure and function. Genome Res. 10(12):1845-1864, 2000. 12

[83] Huntington J.A. and Stein P.E. Structure and properties of ovalbumin. J. Chromatography B: Biomedical Sciences and Applications. 756(1-2): 189-198, 2001. 12, 13

[84] Schechter I. and Berger A. On the size of the active site in proteases. I. Papain. Biochem Biophys Res Commun. 27(2):157-162, 1967. 12

[85] Huntington J.A., Read R.J. and Carrell R.W. Structure of a serpin-protease complex shows inhibition by deformation. Nature. 407(6806): 923-926, 2000. 12

[86] Warrener P., Kim S., Williams S.M.H., Biery M, Gorgon M., Toniatti C., Cleary M.A., Linsley P.S. and Carleton M.C. Synthetic lethality of PARP inhibition in BRCA-network disrupted tumor cells is associated with interferon pathway activation and enhanced by interferon-gamma. Apoptosis. 2012 [Epub ahead of print] 14

[87] Hottiger M.O., Hassa P.O., Luscher B., Schuler and Koch-Nolte F. Toward a unified nomenclature for mammalian ADP-ribosyltransferases. Trends Biochem. Sci. 35: 208-219, 2010. 14, 72

[88] Kotz J. PARP family portraits. Nature Chemical Biology 8: 320, 2012. 14

[89] Yap T.A., Sandhu S.K., Carden C.P. and de Bono J.S. Poly(ADP-Ribose) polymerase (PARP) inhibitors: Exploiting a synthetic lethal strategy in the clinic. CA: A cancer journal for clinicians. 61(1): 31-49, 2011. 14

[90] Ame, J.C., Jacobson, E.L. and Jacobson, M.K. ADP-ribose polymer metabolism. In From DNA Damage and Stress Signalling to Cell Death: Poly ADP-Ribosylation Reactions, G. de Murcia, and S. Shall, eds. (New York: Oxford University Press), pp. 134, 2000. 14

[91] Rolli, V., Ruf, A., Augustin, A., Schulz, G.E., Menissier-de Murcia, J., and de Murcia, G. Poly(ADP-ribose) polymerase: structure and function. In From DNA Damage and Stress Signalling to Cell Death: Poly ADP-Ribosylation Reactions, G. de Murcia, and S. Shall, eds. (New York: Oxford University Press), pp. 35-79, 2000. 14

[92] Kraus W.L. and Lis J.T. PARP Goes Transcription. Cell, Vol. 113: 677-683, 2003. 14

[93] Kim M.Y., Zhang T. and Kraus W.L. Poly(ADP-ribosyl)ation by PARP-1: 'PAR-laying' NAD+ into a nuclear signal. Genes Dev. 19(17): 1951-1967, 2005. 14

[94] Schreiber V., Dantzer F., Ame J.C. and de Murcia G. Poly(ADP-ribose): novel functions for an old molecule. Nature Reviews Molecular Cell Biology 7: 517-528, 2006. 14, 15

[95] Rouleau M., Aubin R.A. and Poirier G.H. Poly(ADP-ribosyl)ated chromatin domains: access granted. J. Cell. Sci. 117: 815-825, 2004. 14

[96] Otto H., Reche P.A., Bazan F., Dittmar K., Haag F. and Koch-Nolte F. In silico characterization of the family of PARP-like poly(ADP-ribosyl)transferases (pARTs). BMC Genomics 6: 139 (2005). 14

[97] Wahlberg E., Karlberg T., Kouznetsolva E., Markova N., Macchiarlulo A., Thorsell A.G., Pol E., Frostell A., Ekblad T., Oncu D., Kull B., Robertson G.M., Pellicciari R., Schuler H. and Weigelt J. Family-wide chemical profiling and structural analysis of PARP and tankyrase inhibitors. Nat. Biotechnol. 30(3): 283-288, 2012. 14, 72

[98] Till S., Diamantara K. and Ladurner A.G. PARP: a transferase by any other name. Nature Struct. Mol. Biol. 15: 1243-1244, 2008. 15, 16

[99] Yu M., Schreek S., Cerni C., Schamberger C., Lesniewicz M., Poreba E., Vervoorts J., Walsemann G., Grtzinger J., Kremmer E., Mehraein Y., Mertsching J., Kraft R., Austen M., Luscher-Firzlaff J. and Luscher B. PARP-10, a novel Myc-interacting protein with poly(ADP-ribose) polymerase activity, inhibit transformation. Oncogene 24: 1982-1993, 2005. 15, 73

[100] Hassa P.O., Haenni S.S., Elser M. and Hottiger M.O. Nuclear ADP-ribosylation reaction in mammalian cells: where are we today and where are we going? Microbiol Mol Biol Rev 70(3): 789-829, 2006. 15

[101] Stambolic V. and Woodgett J.R. Mitogen inactivation of glycogen synthase kinase-3 beta in intact cells via serine 9 phosphorylation. Biochem. J. 303 (Pt 3): 701-704, 1994. 17

[102] Doble B.W. and Woodgett J.R. GSK-3: tricks of the trade for a multitasking kinase, J. Cell Sci. 116: 1175-1186, 2003. 17

[103] Grimes C.A. and Jope R.S. The multifaceted roles of glycogen synthase kinase 3beta in cellular signaling, Prog. Neurobiol. 65: 391-426, 2001. 17

[104] Luo J. Glycogen synthase kinase 3b (GSK3b) in tumorigenesis and cancer chemotherapy. Cancer Letters 273: 194-200, 2009. 17

[105] Hanks S.K., Quinn A.M. and Hunter T. The protein kinase family: conserved features and deduced phylogeny of the catalytic domains. Science 241: 42-52, 1988. 17

[106] ter Haar E., Coll J.T., Austen D.A., Hsiao H.M., Swenson L. and Jain J. Structure of GSK3 reveals a primed phosphorylation mechanism. Nature Struct Biol. 8: 593-596, 2001. 17, 18

[107] Jope R.S. Glycogen synthase kinase-3 (GSK3): inflammation, diseases, and therapeutics. Neurochem Res 32 (4-5): 577-595, 2007. 19

[108] Songin M., Jesko H., Czapski G., Adamczyk A. and Strosznajder R.P. GSK-3B and oxidative stress in aged brain. Role of poly(ADP-ribose) polymerase 1. Folia Neuropathol 45(4): 220-229, 2007. 19

[109] Yeh Y.J., Sbodio J.I. and Chi N.W. Mitotic phosphorylation of tankyrase, a PARP that promotes spindle assembly, by GSK3. Biochem. Biophys. Res. Commun. 350(3): 574-579, 2006. 19, 75

[110] Kalb R. The protean actions of neurotrophins and their receptors on the life and death of neurons. Trends Neurosci, 28: 511, 2005. 19

[111] Reichardt L.F. Neurotrophin-regulated signalling pathways. Philos Trans R Soc Lond B Biol Sci. 361(1473): 1545-1564, 2006. 19

[112] Chao M.V., Rajagopal R. and Lee F.S. Neurotrophin signalling in health and disease. Clin Sci (Lond). 110(2):167-173, 2006. 19

[113] Hock C.H., Heese K., Olivieri G., Hulette C.H., Rosenberg C., Nitsch R.M. and Otten U. Alterations in neurotrophins and neurotrophin receptors in Alzheimer's disease. J Neural Transm Suppl. 59:171-174, 2000. 19

[114] Hefti F. Pharmacology of neurotrophic factors. Annu. Rev. Pharmacol. Toxicol., 37: 239-267, 1997. 19

[115] Hagg T., Manthorpe M., Vahlsing H.L. and Varon S. Delayed treatment with nerve growth factor reverses the apparent loss of cholinergic neurons after acute brain damage. Exp. Neurol. 101: 303-312, 1988. 19

[116] Jonhagen M.E. Nerve growth factor treatment in dementia. Alzheimer Dis Assoc Disord. 14 Suppl 1: 3138, 2000. 19

[117] Tuszynski M.H., Thal L., Pay M., Salmon D.P., U H.S., Bakay R., Patel P., Blesch A., Vahlsing H.L., Ho G., Tong G., Potkin S.G., Fallon J., Hansen L., Mufson E.J., Kordower J.H., Gall C. and Conner J. A phase 1 clinical trial of nerve growth factor gene therapy for Alzheimer disease. Nat. Med. 11: 551-555, 2005. 19

[118] Capsoni S., Ugolini G., Comparini A., Ruberti F., Berardi N. and Cattaneo A. Alzheimer-like neurodegeneration in aged antinerve growth factor transgenic mice. Proc Natl Acad Sci USA 97: 6826-6831, 2000. 19

[119] Capsoni S., Giannotta S. and Cattaneo A. Nerve growth factor and galantamine ameliorate early signs of neurodegeneration in anti-nerve growth factor mice. Proc Natl Acad Sci USA 99: 12432-12437, 2002. 19

[120] De Rosa R., Garcia A.A., Braschi C., Capsoni S., Maffei L., Berardi N. and Cattaneo A. Intranasal administration of nerve growth factor (NGF) rescues recognition mem- ory deficits in AD11 anti-NGF transgenic mice. Proc Natl Acad Sci USA 102: 3811-3816, 2005. 19, 83

[121] Levi-Montalcini R. The nerve growth factor 35 years later. Science 237(4819): 1154-1162, 1987. 19, 83

[122] Dechant G. and Neumann H. Neurotrophins. Adv Exp Med Biol 513: 303-334, 2002 19

[123] Chao M.V. and Hempstead B.L. p75 and trk: a two-receptor system. Trends Neurosci 18: 321-326, 1995. 19

[124] Kaplan D.R. and Miller F.D. Neurotrophin signal transduction in the nervous system.Curr. Opin. Neurobiol. 10: 381-391, 2000. 19, 83

[125] Patapoutian A. and Reichardt L. Trk receptors: mediators of neurotrophin action. Curr. Opin. Neurobiol. 11: 272-280, 2001. 19

[126] Arevalo J.C., Conde B., Hempstead B.L., Chao M.V., Martin-Zanca D. and Perez P. TrkA immunoglobulin-like ligand binding domains inhibit spontaneous activation of the receptor. Mol. Cell Biol. 20: 5908-5916, 2000. 19

[127] Hempstead B. L., Martin-Zanca D., Kaplan D. R., Parada L. F. and Chao, M. V. High-affinity NGF binding requires co-expression of the trk proto-oncogene and the low-affinity NGF receptor. Nature 350: 678683, 1991. 20

[128] Chao M.V. Neurotrophins and their receptors: a convergence point for many signaling pathways. Nat Rev Neurosci 4: 299-309, 2003. 20, 23, 83, 106

[129] Levi-Montalcini R. and Booker B. Destruction of the sympathetic ganglia in mammals by an antiserum to a nerve-growth protein. Proc. Natl. Acad. Sci. USA 46: 384-391, 1960. 22

[130] Yung T.M., Sato S. and Satoh M.S. Poly(ADP-ribosyl)ation as a DNA damage-induced post-translational modification regulating poly(ADP-ribose) polymerase-1-topoisomerase I interaction. J Biol Chem. 279(38): 39686-39696, 2004. 3, 72

[131] Angeletti P.U., Levi-Montalcini R. and Caramia F. Ultrastructural changes in sympa- thetic neurons of newborn and adult mice treated with nerve growth factor. J Ultrastruct Res. 36(1): 2436, 1971. 22

[132] Levi-Montalcini R., Dal Toso R., Della Valle F., Skaper S.D. and Leon A. Update of the NGF saga. J Neurol Sci. 130(2): 119-127, 1995. 22

[133] Fiore M., Chaldakov G.N. and Aloe L. Nerve growth factor as a signaling molecule for nerve cells and also for the neuroendocrine-immune systems. Rev Neurosci. 20(2): 133-145, 2009. 22

[134] Calissano P., Cattaneo A., Biocca S., Aloe L., Mercanti D. and Levi-Montalcini R. The nerve growth factor. established findings and controversial aspects. Exp Cell Res. 154(1): 19, 1984. 22

[135] Stepien H., Lyson K., Stanisz A.M. and Pawlikowski M. The effect of nerve growth factor on DNA synthesis, cyclic AMP and cyclic GMP accumulation by mouse spleen lymphocytes. Int J Immunopharmacol. 113(1): 51-56, 1991. 22

[136] Levi-Montalcini R. and Angeletti P.U. Essential role of the nerve growth factor in the survival and maintenance of dissociated sensory and sympathetic embryonic nerve cells in vitro. Develop. Biol. 7: 653, 1963. 22

[137] Levi-Montalcini R. and Aloe L. Differentiating effects of murine nerve growth factor in the peripheral and central nervous systems of Xenopus laevis tadpoles. Proc. Natl. Acad. Sci. USA 82: 7111-7115, 1985. 22

[138] Zweifel L.S., Kuruvilla R. and Ginty DD. Functions and mechanisms of retrograde neurotrophin signalling. Nat Rev Neurosci 6: 615-625, 2005. 22

[139] Ginty D.D. and Segal RA. Retrograde neurotrophin signaling: Trk-ing along the axon. Curr Opin Neurobiol 12: 268-274, 2002. 22

[140] Miller F.D. and Kaplan D.R. On Trk for retrograde signaling. Neuron 32: 767-770, 2001. 22

[141] Larkfors L., Ebendal T., Whittemore S.R., Persson H., Hoffer B. and Olsen L. Decreased levels of nerve growth factor (NGF) and its messenger RNA in the aged rat brain. Brain Res. 192(427): 55-60, 1987. 22

[142] Bliss T.V. and Collingridge G.L. A synaptic model of memory: long-term potentiation in the hippocampus. Nature 361: 31-39, 1993. 22

[143] Sun W., Sun C., Lin H., Zhao H., Wang J., Ma H., Chen B., Xiao Z. and Dai J. The effect of collagen-binding NGF-beta on the promotion of sciatic nerve regeneration in a rat sciatic nerve crush injury model. Biomaterials 30(27): 4649-4656, 2009. 22

[144] Poduslo J.F. and Curran G.L. Permeability at the blood-brain and blood-nerve barriers of the neurotrophic factors: NGF, CNTF, NT-3, BDNF. Brain Res Mol Brain Res, 36: 280-286, 1996. 22

[145] Hefti F.F., Rosenthal A., Walicke P.A., Wyatt S., Vergara G., Shelton D.L. and Davies A.M. Novel class of pain drugs based on antagonism of NGF. Trends Pharmacol Sci, 27: 85-91, 2005. 22

[146] Saragovi H.U., Greene M.I., Chrusciel R.A. and Kahn M. Loops and secondary structure mimetics: development and applications in basic science and rational drug design. Biotechnology (N Y), 10: 773-778, 1992. 22

[147] Saragovi H.U. and Gehring K. Development of pharmacological agents for targeting neurotrophins and their receptors. Trends Pharmacol Sci 21: 93-98, 2000. 22

[148] Pollack S.J. and Harper S.J. Small molecule trk receptor agonists and other neurotrophic factor mimetics. Curr Drug Target CNS Neurol Disord 1: 59-80, 2002. 22

[149] Massa S.M., Xie Y. and Longo F.M. Alzheimers therapeutics: neurotrophin domain small molecule mimetics. J Mol Neurosci. 20: 323-326, 2003. 22

[150] Berrera M., Cattaneo A. and Carloni P. Molecular Simulation of the Binding of Nerve Growth Factor Peptide Mimics to the Receptor Tyrosine Kinase A. Biophys J. 91(6): 2063-2071, 2006. 22, 25, 91

[151] LeSauteur L., Wei L., Gibbs B.F. and Saragovi H.U. Small peptide mimics of nerve growth factor bind TrkA receptors and affect biological responses. J Biol Chem, 270: 6564-6569, 1995. 23

[152] LeSauteur L., Cheung N.K., Lisbona R. and Saragovi H.U. Small molecule nerve growth factor analogs image receptors in vivo. Nat Biotechnol. 14: 1120-1122, 1996. 23

[153] Xie Y., Tisi M.A., Yeo T.T. and Longo F.M. Nerve growth factor (NGF) loop 4 dimeric mimetics activate ERK and AKT and promote NGF-like neurotrophic effects. J Biol Chem. 275: 29868-29874, 2000. 23

[154] Beglova N., Maliartchouk S., Ekiel I., Zaccaro M.C., Saragovi H.U. and Gehring K. Design and solution structure of functional peptide mimetics of nerve growth factor. J Med Chem, 43: 3530-3540, 2000. 23

[155] Maliartchouk S., Feng Y., Ivanisevic L., Debeir T., Cuello A.C., Burgess K. and Saragovi H.U. Designed peptidomimetic agonistic ligand of TrkA nerve growth factor receptors. Mol Pharmacol. 57: 385-391, 2000. 23

[156] Maliartchouk S., Debeir T., Beglova N., Cuello A.C., Gehring K. and Saragovi H.U. Genuine monovalent ligands of TrkA nerve growth factor receptors reveal a novel pharmacological mechanism of action. J. Biol. Chem. 275: 9946-9956, 2000. 23

[157] Zaccaro M.C., Lee H.B., Pattarawarapan M., Xia Z., Caron A., L-Heureux P.J., Bengio Y., Burgess K. and Saragovi H.U. Selective small molecule peptidomimetic ligands of TrkC and TrkA receptors afford discrete or complete neurotrophic activities. Chem Biol, 12: 1015-1028, 2005. 23

[158] McDonald N.Q., Lapatto R., Murray-Rust J., Gunning J., Wlodawer A. and Blundell T.L. New protein fold revealed by a 2.3-A resolution crystal structure of nerve growth factor. Nature 354 (6352): 411-414, 1991. 23

[159] Bradshaw R.A., Murray-Rust J., Ibanez C.F., McDonald N.Q., Lapatto R. and Blundell T.L. Nerve growth factor: structure/function relationships. Protein Sci. 3(11): 1901-1913, 1994. 23

[160] Timm D.E. and Neet K.E. Equilibrium denaturation studies of mouse beta-nerve growth factor. Protein Sci. 1: 236-244, 1992. 23

[161] Perez P., Coll P.M., Hempstead B.L., Martin-Zanca D. and Chao MV. NGF binding to the trk tyrosine kinase receptor requires the extracellular immunoglobulin-like domains. Mol. Cell. Neurosci. 6(2): 97-105, 1995. 23

[162] Holden P.H., Asopa V., Robertson A.G., Clarke A.R., Tyler S., Bennett G.S., Brain S.D., Wilcock G.K., Allen S.J., Smith S.K. and Dawbarn D. Immunoglobulin-like domains define the nerve growth factor binding site of the TrkA receptor. Nat. Biotechnol. 15(7): 668-672, 1997 23

[163] Covaceuszach S., Cassetta A., Konarev P.V., Gonfloni S., Rudolph R., Svergun D.I., Lamba D. and Cattaneo A. Dissecting NGF interactions with TrkA and p75 receptors by structural and functional studies of an anti-NGF neutralizing antibody. J. Mol. Biol 381(4): 881-896, 2008 23, 91, 97

[164] Zaccaro M.C., Ivanisevic L., Perez P., Meakin S.O. and Saragovi H.U. p75 Co-receptors regulate ligand-dependent and ligand-independent Trk receptor activation, in part by altering Trk docking subdomains. J. Biol. Chem. 276(33): 31023-31029, 2001. 25, 88

[165] Heldin C.H. Dimerization of cell surface receptors in signal transduction. Cell 80(2): 213-223, 1995. 25

[166] Settanni G., Cattaneo A. and Carloni P. Molecular dynamics simulations of the NGF-TrkA domain 5 complex and comparison with biological data. Biophys. J. 84(4): 2282-2292, 2003. 25

[167] Stanzione F., Esposito L., Paladino A., Pedone C., Morelli G. and Vitagliano L. Role of the conformational versatility of the neurotrophin N-terminal regions in their recognition by Trk receptors. Biophys. J. 99(7): 2273-2278, 2010. 25

[168] Cornell W.D., Cieplak P., Bayly C.I., Gould I.R., Merz K.M. Jr, Ferguson D.M., Spellmeyer D.C., Fox T, Caldwell J.W. and Kollman P.A. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. J. Am. Chem. Soc. 117: 5179-5197, 1995. 25

[169] Duan Y., Wu C., Chowdhury S., Lee M.C., Xiong G., Zhang W., Yang R., Cieplak P., Luo R., Lee T., Caldwell J., Wang J. and Kollman P. A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. J. Comput. Chem. 24(16): 1999-2012, 2003. 25

[170] Loytynoja A. and Goldman N. A model of evolution and structure for multiple sequence alignment. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 363(1512): 3913-3919, 2008. 27

[171] McLachlan A.D. Tests for comparing related amino-acid sequences. Cytochrome c and cytochrome c-551. J. Mol. Biol. 61: 409-424, 1971. 28

[172] Dayhoff M.O., Schwartz, R.M. and Orcutt, B.C. A model of evolutionary change in proteins. In "Atlas of Protein Sequence and Structure," Vol. 5, Suppl. 3 (ed. M.O. Dayhoff), 345-352, 1978. 28

[173] Schwartz R.M. and Dayhoff M.O. Matrices for detecting distant relationships. In "Atlas of Protein Sequence and Structure," Vol. 5, Suppl. 3 (ed. M.O. Dayhoff), p. 353-358, 1978. 28

[174] Feng D.F., Johnson M.S. and Doolittle R.F. Aligning amino acid sequences: comparison of commonly used methods. J. Mol. Evol. 21: 112-125, 1984. 28

[175] Fan Y., Wang W., Ma G., Liang L., Shi Q. and Tao S. Patterns of insertion and deletion in Mammalian genomes. Curr Genomics 8(6):370-378, 2007. 28

[176] Lesk A.M., Levitt M. and Clothia C. Alignment of the amino acid sequences of distantly related proteins using variable gap penalties. Protein Eng. 1(1): 77-78, 1986. 28

[177] Zimmerman K.H. Pairwise sequence alignment. AN INTRODUCTION TO PROTEIN INFORMATICS. The Kluwer International Series in Engineering and Computer Science, 749: 47-74, 2003. 28

[178] Philips A., Janies D. and Wheeler W. Multiple Sequence Alignment in Phylogenetic Analysis. Mol. Phylogen. Evol. 16(3): 317-330, 2000. 28

[179] Brudno M., Malde S., Poliakov A., Do C.B., Couronne O., Dubchak I. and Batzoglou S.. Glocal alignment: finding rearrangements during alignment. Bioinformatics 19 Suppl. 1 54: 62, 2003. 28

[180] Errami M., Geourjon C. and Deleage G. Conservation of amino acids into multiple alignments involved in pairwise interactions in three-dimensional protein structures. J. Bioinform. Comput. Biol. 1(3): 505-520, 2003. 28

[181] Lesk A. M. In Introduction to Bioinformatics, First ed.; Oxford University Press: Oxford, UK; Chapter 4, 160-208, 2002. 28

[182] Edgar R.C. and Batzoglou S. Multiple sequence alignment. Curr. Opin. Struct. Biol. 16: 368-373, 2006. 28, 29

[183] Lassmann T. and Sonnhammer E.L.L. Quality assessment of multiple alignment programs. FEBS Letters 529: 126-130, 2002. 28

[184] Lecompte O., Thompson J.D., Plewniak F., Thierry J. and Poch O. Multiple alignment of complete sequences (MACS) in the post-genomic era. Gene. 270(1-2): 17-30, 2001. 28

[185] Feng D.F. and Doolittle R.F. Progressive sequence alignment as a prerequisite to correct phylogenetic trees. J. Mol. Evol. 25(4): 351-360, 1987. 29

[186] Mount D.W. In Bioinformatics. Sequence and Genome Analysis, Second ed.; Cold Spring Harbor Laboratory Press: New York, USA; Chapter 3, 65-115, 2004. 29, 33

[187] Myers G., Selznick S., Zhang Z. and Miller W. Progressive multiple alignment with constraints. J Comput Biol. 3(4): 563-572, 1996. 29

[188] Bray N. and Pachter L. MAVID: constrained ancestral alignment of multiple sequences. Genome Res. 14(4):693-699, 2004. 29

[189] Loytynoja A. and Milinkovitch M.C. A hidden Markov model for progressive multiple alignment. Bioinformatics 19(12): 1505-1513, 2003. 29

[190] Morgenstern B., Werner N., Prohaska S.J., Steinkamp R., Schneider I., Subramanian A.R., Stadler P.F. and Weyer-Menkhoff J. Multiple sequence alignment with user-defined constraints at GOBICS. Bioinformatics. 21(7):1271-1273, 2004. 29

[191] Thompson J., Higgins D. and Gibson T. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucl. Ac. Res. 22: 4673-4680, 1994. 29, 36

[192] Li K.B. ClustalW-MPI: ClustalW analysis using distributed and parallel computing. Bioinformatics 19(12): 1585-1586, 2003. 29

[193] Gibbs A and McIntyre G. The Diagram, a Method for Comparing Sequences. Its use with Amino Acid and Nucleotide Sequences. Eur. J. Biochem. 16, 1-11, 1970. 29

[194] Needleman S.B. and Wunsch C.D. A general method applicable to the search for similarities in the amino acid sequence of two proteins. J. Mol. Biol. 48(3): 443-453, 1970. 29, 30

REFERENCES

[195] Smith, Temple F.; and Waterman, Michael S. Identification of Common Molecular Subsequences. J. Mol. Biol. 147: 195-197, 1981. 29, 33

[196] Pearson W. and Lipman D. Improved tools for biological sequence comparison. Proc. Natl. Acad. Sci. U.S.A. 85: 2444-2448, 1988. 29, 33

[197] Altschul S.F., Gish W., Miller W., Myers E. and Lipman D. Basic Local Alignment Search Tool. J. Mol. Biol. 215: 403, 1990. 29, 33, 36

[198] Brown T.A. Genomes: 2nd edition. Oxford: Wiley-Liss, 2002. 30

[199] Lobo, I. Basic Local Alignment Search Tool (BLAST). Nature Education 1(1), 2008. 31

[200] Eddy S.R. What is dynamic programming? Nat. Biotec. 22(7): 909-910, 2004. 31

[201] Drmanac R., Labat I. and Crkvenjakov R. An algorithm for the DNA sequence generation from k-tuple word contents of the minimal number of random fragments. J. Biomol. Struct. Dyn. 8(5): 1085-1102, 1991. 31

[202] Capra J.A. and Singh M. Predicting functionally important residues from sequence conservation. Bioinformatics. 23(15):1875-1882, 2007. 32

[203] Dayhoff M. and Schwartz R. Matrices for detecting distant relationship. Atlas of Protein Sequences. 353-358, 1979. 32

[204] Henikoff S. and Henikoff J.G. Amino acid substitution matrices from protein blocks. Proc.Natl.Acad.Sci. USA, 89(22): 10915-10919, 1992. 32

[205] Plewniak F. Database similarity searches. Methods Mol Biol. 484: 361-378, 2008. 33

[206] Bader D.A. and Sachdeva V. An open benchmark suite for evaluating computer architecture on bioinformatic and life science applications. SPEC Benchmark Workshop, 2006. 33

[207] Marti-Renom M.A., Stuart A.C., Fiser A., Sanchez R., Melo F. and Sali A. Comparative protein structure modeling of genes and genomes. Ann. Rev. Biophys. Biomol. Struct. 29: 291-325, 2000. 34, 35

[208] Flohil J.A., Vriend G. and Berendsen H.J.C. Completion and Refinement of 3-D Homology Models with Restricted Molecular Dynamics: Application to Targets 47, 58, and 111 in the CASP Modeling Competition and Posterior Analysis. Proteins 48: 593-604, 2002. 34

[209] Epstein C.J., Goldberger R.F. and Anfinsen C.B. The genetic control of tertiary protein structure: studies with model systems. Cold Spring Harb. Symp. Quant. Biol. 28: 439-449, 1963. 34

[210] Chothia C. and Lesk A.M. The relation between the divergence of sequence and structure in proteins. Embo J. 5(4): 823-826, 1986. 34, 37

[211] Sander C. and Schneider R. Database of homology-derived protein structures and the structural meaning of sequence alignment. Proteins. 9(1): 56-68, 1991. 34

[212] Rost B. Twilight zone of protein sequence alignments. Protein Eng. 12(2): 85-94, 1999. 34

[213] Sternberg M.J.E. Protein structure prediction: Principles and approaches In Protein structure prediction 130. Oxford University Press, Oxford, UK, 1996. 35

[214] Sanchez R. and Sali A. Advances in comparative protein structure modeling. Curr. Opin. Struct. Biol. 7: 206-214, 1997. 35

[215] Eswar N., Webb B., Marti Renom M.A., Madhusudhan M.S., Eramian D., Shen M., Pieper U. and Sali A. Comparative Protein Structure Modeling Using MODELLER. Curr. Protocols in Prot. Sci. 50: 2.9.12.9.31, 2007. 35

[216] Bernstein F.C., Koetzle T.F., Williams G.J., Meyer E.F., Brice M.D., Rodgers J.R., Kennard O., Shimanouchi T. and Tasumi M. The Protein Data Bank: a computer-based archival file for macromolecular structures. J. Mol. Biol. 112(3): 535-542, 1977. 35

[217] Berman H.M., Westbrook J., Feng Z., Gilliland G., Bhat T.N., Weissig H., Shindyalov I.N. and Bourne P.E. The Protein Data Bank. Nucleic Acid Res 128(1): 235-242, 2000. 35, 76

[218] Pearson W.R. Rapid and sensitive sequence comparison with FASTP and FASTA. Methods Enzymol. 183: 63-98, 1990. 36

[219] Krieger E., Nabuurs S.B. and Vriend G. in: Structural Bioinformatics. Wiley-Liss, Inc., 2003. 36, 37

[220] Higgins D.G. and Sharp P.M. CLUSTAL: a package for performing multiple sequence alignment on a microcomputer. Gene 73: 237-244, 1988. 36

[221] Sali A. and Blundell T.L. Comparative protein modelling by satisfaction of spatial restraints. J Mol Biol. 234(3): 779-815, 1993. 36, 60

[222] Guex N. and Peitsch M.C. SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling. Electrophoresis 18, 2714-2723, 1997. 36, 61

[223] Bower M.J., Cohen F.E., and Dunbrack R.L. Prediction of protein side-chain rotamers from a backbone-dependent rotamer library: A new homology modeling tool. J. Mol. Biol. 267: 1268-1282, 1997. 36

[224] Bates P.A., Kelley L.A., MacCallum R.M. and Sternberg M.J. Enhancement of protein modeling by human intervention in applying the automatic programs 3D-JIGSAW and 3D-PSSM. Proteins Suppl 5: 39-46, 2001. 36

[225] Schwede T., Kopp J., Guex N. and Peitsch M.C. SWISS-MODEL: An automated protein homology-modeling server. Nucleic Acids Res 31(13): 3381-3385, 2003. 36

[226] Kaistha S.D. and Sinha R. Homology modeling of phosphoryl thymidine kinase of enterohemorrhagic Escherichia coli OH: 157. Bioinformation. 3(6): 240-243, 2009. 36

[227] MacKerell A.D., Bashford D., Bellott M., Dunbrack R., Evanseck J., Field M., Fischer S., Gao J., Guo H., Ha S., Joseph-McCarthy D., Kuchnir L., Kuczera D., Lau F.T.K., Mattos C., Michnick S., Ngo T., Nguyen D.T., Prodhom B., Reiher W.E., Roux B., Schlenkrich M., Smith J.C., Stote R., Straub J., Watanabe M., Wiorkiewica-Kuczera J., Yin D. and Karplus M. All-atom empirical potential for molecular modeling and dynamics studies of proteins. J. Phys. Chem. B. 102: 3586-3616, 1998. 37

[228] Wallner B. and Elofsson A. All are not equal: A benchmark of different homology modeling programs. Protein Sci. 14(5): 1315-1327, 2005. 37

[229] Fiser A., Do R.K. and Sali A. Modeling of loops in protein structures. Protein Sci. 9(9): 1753-1773, 2000. 37

[230] Raval A., Piana S., Eastwood M.P., Dror R.O. and Shaw DE. Refinement of protein structure homology models via long, all-atom molecular dynamics simulations. Proteins. doi: 10.1002/prot.24098, 2012. 37

[231] Xiang Z. Advances in Homology Protein Structure Modeling. Curr. Prot. Pept. Sci. 7(3): 217-227, 2006. 37

[232] Lee M.R., Baker D. and Kollman P.A. 2.1 and 1.8 A average C(alpha) RMSD structure predictions on two small proteins, HP-36 and s15. J Am Chem Soc. 123: 10401046, 2001. 37

[233] Fan H. and Mark A.E. Refinement of homology-based protein structures by molecular dynamics simulation techniques. Protein Sci. 13(1): 211-220, 2004. 37

[234] Qian B., Ortiz A.R. and Baker D. Improvement of comparative model accuracy by free-energy optimization along principal components of natural structural variation. Proc Natl Acad Sci. 101(43): 15346-15351, 2004. 37

[235] Luthy R., Bowie J.U. and Eisenberg D.. Assessment of protein models with three-dimensional profiles. Nature 356, 83-85, 1992. 37

[236] Laskowski R.A., Rullmann J.A., MacArthur M.W., Kaptein R. and Thornton J.M. AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. J Biomol NMR 8: 477-486, 1996. 37

[237] Oldfield T.J. SQUID: a program for the analysis and display of data from crystallography and molecular dynamics. J. Mol. Graph. 10(4): 247-252, 1992. 37

[238] Hooft R.W.W., Vriend G., Sander C. and Abola E.E. Errors in protein structures. Nature 381, 272-272, 1996. 37

[239] Walker J.M. The proteomics protocol handbook. Humana Press Inc., 2005. 37

[240] Word J.M., Bateman R.C., Presley B.K., Lovell S.C. and Richardson D.C. Exploring steric constraints on protein mutations using MAGE/PROBE. Protein Sci 9: 2251-2259, 2000. 37

[241] Bhattacharya A., Wunderlich Z., Monleon D., Tejero R. and Montelione G. Assessing model accuracy using the homology modeling automatically software. Proteins 70(1): 105-118, 2008. 38

[242] Sippl MJ, Calculation of conformational ensembles from potentials of mean force. An approach to the knowledge-based prediction of local structures in globular proteins. J. Mol. Biol. 213: 859-883, 1990. 38

[243] Eisenberg D., Luthy R. and Bowie J.U. VERIFY3D: assessment of protein models with three-dimensional profiles. Methods Enzymol. 277: 396-404, 1997. 38

[244] Melo F. and Feytmans E. Assessing protein structures with a non-local atomic interaction energy. J. Mol. Biol. 277(5): 1141-1152, 1998. 38

[245] Di Luccio E. and Koehl P. A quality metric for homology modeling: the H-factor. BMC Bioinformatics 12: 48, 2011. 38

[246] Kryshtafovych A., Fidelis K. and Moult J. CASP9 results compared to those of previous casp experiments. Proteins 79 Suppl 10: 196-207, 2011. 38

[247] Venselaar H., Joosten R.P., Coos B.V., Baakman A.B., Hekkelman M.L., Krieger E., Vriend G. Homology modelling and spectroscopy, a never-ending love story. Eur Biophys J. 39(4): 551-563, 2010. 38

[248] Ehrlich L.P., Nilges M. and Wade R.C. The impact of protein flexibility on protein-protein docking. Proteins. 58(1): 126-133, 2005. 38

[249] Zacharias M. Accounting for conformational changes during protein-protein docking. Curr. Opin. Struct. Biol. 20(2): 180-186, 2010. 39

[250] Andrusier N., Mashiach E., Nussinov R. and Wolfson, H. J. Principles of flexible protein-protein docking. Proteins 73: 271-289, 2008. 39

[251] Tsai C.J., Ma B., Sham Y.Y., Kumar S. and Nussinov R. Structured disorder and conformational selection. Proteins. 44(4): 418-427, 2001. 39

[252] Goh C.S., Milburn D. and Gerstein M. Conformational changes associated with protein-protein interactions. Curr. Opin. Chem. Biol. 14: 16, 2004. 39

[253] Boehr D.D. and Wright P.E. How do proteins interact? Science 320(5882): 1429-1430, 2008. 39

[254] De Vries S.J., van Dijk A.D.J, Krzeminski M., van Dijk M., Thureau A., Hsu V., Wassenaar T., Bonvin A.M.J.J. HADDOCK versus HADDOCK: New features and performance of HADDOCK2.0 on the CAPRI targets. Proteins 69: 726-733, 2007. 39

[255] De Vries S.J., van Dijk M. and Bonvin A.M. J.J. The HADDOCK web server for data-driven biomolecular docking. Nat. Protocols 5: 883-897, 2010. 39, 77

[256] Jorgensen W.L., Chandrasekhar J., Madura J.D., Impey R.W. and Klein M. L. Comparison of simple potential functions for simulating liquid water. J. Chem. Phys. 79: 926, 1983. 40, 61, 85

[257] Jorgensen W.L. and Tirado-Rives J. The OPLS Force Field for Proteins. Energy Minimizations for Crystals of Cyclic Peptides and Crambin. J. Am. Chem. Soc. 110(6): 1657-1666, 1988. 40

[258] Bujnicki J.M. Predictions of protein structures, function and interactions. Wiley 2009. 41

[259] Krissinel E. and Henrick K. Inference of macromolecular assemblies from crystalline state. J. Mol. Biol. 372: 774-797, 2007. 41

[260] Horovitz A. and Fersht A.R. Strategy for analysing the co-operativity of intramolecular interactions in peptides and proteins. J Mol Biol. 214(3): 613-617, 1990. 42

[261] Fersht A. The hydrogen bond in molecular recognition. Trends Biochem. Sci., 12: 3214-3219, 1987. 42

[262] Betz S. Disulphide bonds and the stability of globular proteins. Protein Sci. 2: 1551-1558, 1993. 42

[263] Kruger D.M. and Gohlke H. DrugScorePPI webserver: fast and accurate in silico alanine scanning for scoring proteinprotein interactions. Nucl. Acids Res. 38(suppl 2): W480-W486, 2010. 42

[264] Kortemme T., Kim D.E. and Baker D. Computational alanine scanning of protein-protein interfaces. Sci STKE. 2004(219):pl2, 2004. 42, 58, 59, 61, 64, 100, 102

[265] Kortemme T., Morozov A. V. and Baker D. An orientation-dependent hydrogen bonding potential improves prediction of specificity and structure for proteins and protein-protein complexes. J Mol Biol 326: 1239-1259, 2003. 42

[266] Lazaridis T. and Karplus M. Effective energy function for proteins in solution. Proteins 35(2): 133-152, 1999. 42

[267] Thorn K.S. and Boden A.A. ASEdb: a database of alanine mutations and their effects on the free energy of binding in protein interactions. Bioinformatics 17(3): 284-285, 2001. 42

[268] Vajda S. and Kozakov D. Convergence and Combination of Methods in Protein-Protein Docking. Curr Opin Struct Biol. 19(2): 164-170, 2009. 42

[269] Szilagyi A., Grimm V., Arakaki A.K. and Skolnick J. Prediction of physical protein-protein interactions. Phys Biol. 2(2): S1-16, 2005. 42

[270] Wells J A Systematic mutational analyses of protein-protein interfaces. Methods Enzymol. 202: 330-341, 1991. 42

[271] Phizicky E., Bastiaens P.I., Zhu H., Snyder M. and Fields S. Protein analysis on a proteomic scale. Nature 422: 208-211, 2003. 42

[272] Rappsilber J., Siniossoglou S., Hurt E.C. and Mann M. A generic strategy to analyze the spatial organization of multi-protein complexes by cross-linking and mass spectrometry. Anal. Chem. 72: 267-275, 2000. 43

[273] Melcher K. New chemical crosslinking methods for the identification of transient protein-protein interactions with multiprotein complexes. Curr. Protein Pept. Sci. 5: 287-296, 2004. 43

[274] Takahashi H., Nakanishi T., Kami K., Arata Y. and Shimada I. A novel NMR method for determining the interfaces of large protein-protein complexes. Nat. Struct. Biol. 7: 220-223, 2000. 43

[275] Morrison J., Yang J. C., Stewart M. and Neuhaus D. Solution NMR study of the interaction between NTF2 and nucleoporin FxFG repeats. J. Mol. Biol. 333: 587-603, 2003. 43

[276] Takeuchi K., Takahashi H., Sugai M., Iwai H., Kohno T., Sekimizu K., Natori S. and Shimada I. Channel-forming membrane permeabilization by an antibacterial protein, sapecin: determination of membrane-buried and oligomerization surfaces by NMR. J. Biol. Chem. 279: 4981-4987, 2004. 43

[277] Karplus M. and McCammon A. Molecular dynamics simulations of biomolecules. Nat. Struct. Biol. 9(9): 646-652, 2002. 43

[278] Frenkel, D. and Smit B. Understanding Molecular Simulation. From algorithms to applications, 2nd ed.; Academic Press: San Diego, USA, 2002. 44, 46

[279] Phillips J.C., Braun R., Wang W., Gumbart J., Tajkhorshid E., Villa E., Chipot C., Skeel R.D., Kale L., and Schulten K. Scalable molecular dynamics with namd. J. Comput. Chem., 26: 1781-1802, 2005. 45, 54, 62, 85

[280] Feenstra K.A., Hess B. and Berendsen H.J.C. Improving efficiency of large time-scale molecular dynamics simulations of hydrogen-rich systems. J. Comput. Theor. Chem. 20(8): 786-798, 1999. 45

[281] Hess B., Bekker H., Berendsen H.J.C., Fraaije, G.E.M. Lincs: A linear constraint solver for molecular simulations. J. Comput. Chem. 18: 1463, 1997. 45

[282] Verlet L. Computer experiments on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules. Phys. Rev. 159: 98-103, 1967. 45

[283] Hockney R.W. The potential calculation and some applications. Methods Comp. Phys. 9: 135-211, 1974. 45

[284] Ponder J.W. and Case D.A. Force field for protein simulations. Adv. Prot. Chem. 66: 27-85, 2003. 47

[285] Hornak V., Abel R., Okur A., Strockbine B., Roltberg A. and Simmerling C. Comparison of multiple Amber force fields and development of improved protein backbone parameters. Proteins 65(3): 712-725, 2006. 47, 61, 85

[286] Gonzalez M.A. Force fields and molecular dynamics simulations. Collection SFN 12: 169-200, 2011. 47, 49

[287] Bayly C.I., Cieplak P., Cornell W. and Kollman P.A. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. J. Phys. Chem. 97(40): 10269-10280, 1993. 48

[288] Allen M.P. and Tildesley D.J. Computer simulation of liquids. Oxford University Press, 1989. 50

[289] York D.M., Darden T.A. and Pedersen L.G. The effect of long-range electrostatic interactions in simulation of macromolecular crystals: a comparison of the Ewald and truncated list method. J. Chem. Phys. 99: 8345-8348, 1993. 50

[290] Ewald P. Die Berechnung optischer und elektrostatischer Gitterpotentiale. Ann. Phys. 369: 253-287, 1921. 50

[291] Sagui C. and Darden T.A. Molecular dynamics simulations of biomolecules: long range electrostatic effect. Annu. Rev. Biophys. Biomol. Struct. 28: 155-179, 1999. 51

[292] Nose S. Molecular dynamics method for simulation in the canonical ensemble. Mol. Physics, 52:255-268, 1984. 52, 53

[293] Evans D.J. and Holian B.L. The Nose-Hoover thermostat. J. Chem. Phys., 83:4069-4074, 1985. 52, 53

[294] Adelman S.A and Doll J.D. Generalized langevin equation approach for atom-solid-surface scattering - general formulation for classical scattering off harmonic solids. J. Chem. Phys., 64: 2375-2388, 1976. 53, 61

[295] Martyna G.J. Remarks on "Constant-temperature molecular dynamics with momentum conservation". Phys Rev E Stat Phys Plasmas Fluids Relat Interdiscip Topics. 50(4): 3234-3236, 1994. 54

[296] Feller S.E., Zhang Y., Pastor R.W. and Brooks B.R. Constant pressure molecular dynamics simulation: The Langevin piston method. J. Chem. Phys. 103: 4613, 1995. 54, 61

[297] Carter E., Ciccotti G., Hynes J. and Kapral R. Constrained reaction coordinate dynamics for the simulation of rare events. Chem. Phys. Lett. 156: 472, 1989. 55

[298] Sprik M. and Ciccotti G. Free energy from constrained molecular dynamics. J. Chem. Phys. 109: 7737, 1998. 55

[299] Patey G.N. and Valleau J.P. Monte-carlo method for obtaining interionic potential of mean force in ionic solution. J. Chem. Phys. 63: 2334-2339, 1975. 55

[300] Gullingsrud J., Braun R. and Schulten K. Reconstructing potentials of mean force through time series analysis of steered molecular dynamics simulations. J. Comp. Phys. 151: 190-211, 1999. 55

[301] Ferrario M., Ciccotti G. and Binder K. Computer Simulations in Condensed Matter Systems: From Materials to Chemical Biology Volume 1. Springer Berlin Heidelberg: 2006. ISBN 9783540352730. 55

[302] Laio A. and Parrinello M. Escaping free energy minima. Proc. Natl. Acad. Sci. USA 99: 12562-12566, 2002. 55, 85

[303] Daura X., Van Gunsteren W.F. and Mark A.E. Folding-unfolding thermodynamics of a beta-heptapeptide from equilibrium simulations. Angew. Chem. Int. Ed. 34: 269, 1999. 55

[304] van Gunsteren W.F., Bakowies D., Baron R., Chandrasekhar I., Christen M., Daura X., Gee P., Geerke D.P., Glattli A., Hunenberger P.H., Kastenholz M.A., Oostenbrink C., Schenk M., Trzesniak D., van der Vegt N.F. and Yu H.B. Angew Chem Int Ed Engl. 45(25): 4064-4092, 2006. 55

[305] Grubmuller H. Predicting slow structural transitions in macromolecular systems: Conformational flooding. Phys Rev E Stat Phys Plasmas Fluids Relat Interdiscip Topics. 52(3): 2893-2906, 1995. 55

[306] Grubmuller H., Heymann B. and Tavan P. Ligand binding: molecular mechanics calculation of the streptavidin-biotin rupture force. Science 271(5251): 997-999, 1996. 55

[307] Barducci, A., Bussi, G. and Parrinello, M. Well-tempered metadynamics: a smoothly converging and tunable free-energy method. Phys. Rev. Lett. 100(2), 020603 (2008) 56

[308] Singh S., Chiu C. and de Pablo JJ. Flux tempered metadynamics. J. Stat. Phys. 2011, 932-945.

[309] Barducci A., Bonomi M. and Parrinello M. Metadynamics. WIREs Comput Mol Sci 1(5): 826-843, 2011. 56

[310] Lord S.J., Rajotte R.V., Korbutt G.S. and Bleackley R.C. Granzyme B: a natural born killer. Immunol. Rev. 193:31-38, 2003. 58, 59

[311] Kuwana T., Mackey M.R., Perkins G., Ellisman M.H., Latterich M., Schneiter R., Green D.R., Newmeyer D.D. Bid, Bax, and lipids cooperate to form supramolecular openings in the outer mitochondrial membrane. Cell 111(3):331-342, 2002. 58

[312] Waterhouse N.J., Sedelies K.A., Browne K.A., Wowk M.E., Newbold A., Sutton V.R., Clarke C.J., Oliaro J., Lindemann R.K., Bird P.I., Johnstone R.W. and Trapani J.A. A central role for Bid in granzyme B-induced apoptosis. J. Biol. Chem. 280(6): 4476-4482, 2005. 58

[313] Waterhouse N.J., Sedelies K.A., Sutton V.R., Pinkoski M.J., Thia K.Y., Johnstone R., Bird P.I., Green D.R. and Trapani J.A. Functional dissociation of DeltaPsim and cytochrome c release defines the contribution of mitochondria upstream of caspase activation during granzyme B-induced apoptosis. Cell. Death Differ. 13(4): 607-618, 2006. 58

[314] Waterhouse N.J., Sedelies K.A. and Trapani J.A. Role of Bid-induced mitochondrial outer membrane permeabilization in granzyme B-induced apoptosis. Immunol. Cell. Biol. 84(1): 72-78, 2006. 58

[315] Jiang X. and Wang X. Cytochrome C-mediated apoptosis. Annu. Rev. Biochem. 73: 87-106, 2004. 58

[316] Kroemer G., Galluzzi L. and Brenner C. Mitochondrial membrane permeabilization in cell death. Physiol. Rev. 87(1): 99-163, 2007 58

[317] Kurschus F.C., Fellows E., Stegmann E. and Jenne D.E. Granzyme B delivery via perforin is restricted by size, but not by heparan sulfate-dependent endocytosis. Proc. Natl. Acad. Sci. U S A 105(37): 13799-13804, 2008. 58

[318] de Koning P.J., Kummer J.A., de Poot S.A., Quadir R., Broekhuizen R., McGettrick A.F., Higgins W.J., Devreese B., Worrall D.M. and Bovenschen N. Intracellular serine protease inhibitor SERPINB4 inhibits granzyme M-induced cell death. PLoS One 6(8):e22645, 2011. 58

[319] Ye S., Cech A.L., Belmares R., Bergstrom R.C., Tong Y., Corey D.R., Kanost M.R. and Goldsmith E.J. The structure of a Michaelis serpin-protease complex. Nat. Struct. Biol. 8(11):979-983, 2001. 58, 59, 101, 102

[320] Hedstrom L. Serine protease mechanism and specificity. Chem. Rev. 102(12): 4501-4524, 2002. 58

[321] Silverman G.A., Bird P.I., Carrell R.W., Church F.C., Coughlin P.B., Gettins P.G., Irving J.A., Lomas D.A., Luke C.J., Moyer R.W., Pemberton P.A., Remold-O'Donnell E., Salvesen G.S., Travis J., and Whisstock J.C. The serpins are an expanding superfamily of structurally similar but functionally diverse proteins. Evolution, mechanism of inhibition, novel functions, and a revised nomenclature. J. Biol. Chem. 276(36): 33293-33296, 2001. 58, 101

[322] Lise S., Archambeau C., Pontil M. and Jones D.T. Prediction of hot spot residues at protein-protein interfaces by combining machine learning and energy-based methods. BMC Bioinformatics 10: 365, 2009 59

[323] Ruggles S.W., Fletterick R.J. and Craik C.S. Characterization of structural determinants of granzyme B reveals potent mediators of extended substrate specificity. J. Biol. Chem. 279(29): 30751-30759, 2004. 59

[324] Rawlings N.D. and Barrett A.J. Families of serine peptidases. Methods Enzymol. 244: 19-61, 1994. 59

[325] Sun J., Ooms L., Bird C.H., Sutton V.R., Trapani J.A. and Bird PI. A new family of 10 murine ovalbumin serpins includes two homologs of proteinase inhibitor 8 and two homologs of the granzyme B inhibitor (proteinase inhibitor 9). J. Biol. Chem. 272(24):15434-15441, 1997. 60

[326] Laskowski R.A., Macarthur M.W., Moss D.S. and Thornton J.M. Procheck - a Program to Check the Stereochemical Quality of Protein Structures. J. Appl. Crystallogr. 26: 283-291, 1993 60

[327] Shen M.Y. and Sali A. Statistical potential for assessment and prediction of protein structures. Prot. Sci. 15(11): 2507-2524, 2006 60

[328] Aaqvist J. Ion-Water Interaction Potentials derived from Free Energy Perturbation Simulations. J. Phys. Chem. 94: 8021-8024, 1990. 61

[329] Ryckaert J-P., Ciccotti G. and Berendsen H.J.C. Numerical integration of the Cartesian Equations of Motions of a System with Constraints: Molecular Dynamics of n-Alkanes. J. Comp. Phys. 23: 327-341, 1977. 61

[330] Darden T. YD and Pedersen L. Particle mesh Ewald: an N-log(N) method for Ewald sums in large systems. J. Chem. Phys. 98: 10089-10093, 1993. 61

[331] Moreira I.S., Fernandes P.A. and Ramos M.J. Computational alanine scanning mutagenesis–an improved methodological approach. J. Comp. Chem. 28(3): 644-654, 2007. 61

[332] Benedix A., Becker C.M., de Groot B.L., Caflisch A. and Bockmann R.A. Predicting free energy changes using structural ensembles. Nat. Methods 6(1): 3-4, 2009. 61

[333] Kollman P.A., Massova I., Reyes C., Kuhn B., Huo S., Chong L., Lee M., Lee T., Duan Y., Wang W., Donini O., Cieplak P., Srinivasan J., Case D.A. and Cheatham TE, 3rd. Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. Acc. Chem. Res. 33(12):889-897, 2000. 61

[334] Micheletti C., Seno F. and Maritan A. Recurrent oligomers in proteins: an optimal scheme reconciling accurate and concise backbone representations in automated folding and design studies. Proteins 40(4): 662-674, 2000. 62

[335] Stocker M., Tur M.K., Sasse S., Krussmann A., Barth S., Engert A. Secretion of functional anti-CD30-angiogenin immunotoxins into the supernatant of transfected 293T-cells. Protein Expr. Purif. 28(2): 211-219, 2003 62

[336] Pham C.T., Thomas D.A., Mercer J.D. and Ley T.J. Production of fully active recombinant murine granzyme B in yeast. J Biol Chem 273(3): 1629-1633, 1998. 62

[337] Futamura N., Aluru S., Ranjan D. and Hariharan B. Efficient parallel algorithms for solvent accessible surface area of proteins. IEEE Trans Parallel Distrib Syst 13(6): 544-555, 2002. 63

[338] Sun J., Whisstock J.C., Harriott P., Walker B., Novak A., Thompson P.E., Smith A.I. and Bird P.I. Importance of the P4' residue in human granzyme B inhibitors and substrates revealed by scanning mutagenesis of the proteinase inhibitor 9 reactive center loop. J Biol Chem 276(18): 15177-15184, 2001 68

[339] Hanahan D. and Weinberg R.A. The hallmarks of cancer. Cell. 100(1): 57-70, 2000. 71

[340] Mangerich A. and Burkle A. How to kill tumor cells with inhibitors of poly(ADP-ribosyl)ation. Int. J. Cancer. 128(2): 251-265, 2011. 71

[341] Madhusudan S. and Middleton M.R. The emerging role of DNA repair proteins as predictive, prognostic and therapeutic targets in cancer. Cancer Treat Rev. 31(8):603-617, 2005. 71

[342] Helleday T., Petermann E., Lundin C., Hodgson B. and Sharma R.A. DNA repair pathways as targets for cancer therapy. Nat Rev Cancer. 8(3): 193-204, 2008. 71

[343] Helleday T. Amplifying tumour-specific replication lesions by DNA repair inhibitors - a new era in targeted cancer therapy. Eur J. Cancer. 44(7): 921-927, 2008. 71

[344] Yelamos J., Farres J., Llacuna L., Ampurdanes C. and Martin-Caballero J. PARP-1 and PARP-2: New players in tumour development. Am J Cancer Res. 1(3):328-346, 2011. 72

[345] Ame J.C., Spenlehauer C. and de Murcia G. The PARP superfamily. BioEssays 26: 882-893, 2004. 72

[346] Gibson B.A. and Kraus W.L. New insights into the molecular and cellular functions of poly(ADP-ribose) and PARPs. Nat Rev Mol Cell Biol. 13(7):411-424, 2012. 72

[347] Hothorn M., Neumann H., Lenherr E.D., Wehner M., Rybin V., Hassa P.O., Uttenweiler A., Reinhardt M., Schmidt A., Seiler J., Ladurner A.G., Herrmann C., Scheffzek K. and Mayer A. Catalytic core of a membrane-associated eukaryotic polyphosphate polymerase. Science. 324(5926): 513-516, 2009. 72

[348] Luo J. Glycogen synthase kinase 3β (GSK3β) in tumorigenesis and cancer chemotherapy. Cancer Lett. 273: 194200, 2009. 75

[349] Hur E.M. and Zhou F.Q. GSK3 signalling in neural development. Nat Rev Neurosci. 11(8): 539-551, 2010. 75

[350] Luykx J.J., Boks M.P., Terwindt A.P., Bakker S., Kahn R.S. and Ophoff R.A. The involvement of GSK3 in bipolar disorder: integrating evidence from multiple types of genetic studies. Eur Neuropsychopharmacol. 20(6): 357-368, 2010. 75

[351] Hernandez F., Nido J.D., Avila J. and Villanueva N. GSK3 inhibitors and disease. Mini Rev Med Chem. 9(9): 1024-1029, 2009. 75

[352] Kwok J.B., Hallupp M., Loy C.T., Chan D.K., Woo J., Mellick G.D., Buchanan D.D., Silburn P.A., Halliday G.M. and Schofield P.R. GSK3 polymorphisms alter transcription and splicing in Parkinson's disease. Ann. Neurol. 58(6): 829-839, 2005. 75

[353] Corda D. and Di Girolamo M. Functional aspects of protein mono-ADP- ribosylation. EMBO J. 22(9): 1953-1958, 2003. 75, 81

[354] Dajani R., Fraser E., Roe S.M., Young N., Good V., Dale T.C. and Pearl L.H. Crystal structure of glycogen synthase kinase 3 beta: structural basis for phosphate-primed substrate specificity and autoinhibition. Cell. 105(6): 721-732, 2001. 75

[355] Berg S., Bergh M., Hellberg S., Hogdin K., Lo-Alfredsson Y., Soderman P., von Berg S., Weigelt T., Ormo M., Xue Y., Tucker J., Neelissen J., Jerning E., Nilsson Y. and Bhat R. Discovery of Novel Potent and Highly Selective Glycogen Synthase Kinase-3 (GSK3) Inhibitors for Alzheimer's Disease: Design, Synthesis, and Characterization of Pyrazines. J. Med. Chem. DOI: 10.1021/jm201724m, 2012. 75

[356] Bax B., Carter P.S., Lewis C., Guy A.R., Bridges A., Tanner R., Pettman G., Mannix C., Culbert A.A., Brown M.J., Smith D.G. and Reith A.D. The structure of phosphorylated GSK-3beta complexed with a peptide, FRATtide, that inhibits beta-catenin phosphorylation. Structure. 9(12): 1143-1152, 2001. 75

[357] Lee Y., Satheesan Babu C., Chen Y., Milcic M., Qu Y. and Lim C. Conserved structural motif for recognizing nicotinamide adenine dinucleotide in poly(ADP-ribose) polymerases and ADP-ribosylating toxins: implications for structure-based drug design. J. Med. Chem. 53: 4038-4049, 2010. 76, 79

[358] Karaca E., Melquiond A.S., de Vries S.J., Kastritis P.L. and Bonvin A.M. Building macromolecular assemblies by information-driven docking: introducing the HADDOCK multibody docking server. Mol Cell Proteomics. 9(8): 1784-1794, 2010. 77

[359] Lange A., Hoeller D., Wienk H., Marcillat O., Lancelin J.M. and Walker O. NMR Reveals a Different Mode of Binding of the Stam2 VHS Domain to Ubiquitin and Diubiquitin. Biochemistry 50(1): 48-62, 2011. 77

[360] Fushman D. and Walker O. Exploring linkage dependence of polyubiquitin conformations using molecular modeling. J Mol Biol. 395(4): 803, 2010. 77

[361] Jorgensen R., Wang Y., Visschedyk D. and Merrill A.R. The nature and character of the transition state for the ADP-ribosyltransferase reaction. EMBO Rep. 9(8): 802-809, 2008. 79

[362] Bell C.E. and Eisenberg D. Crystal structure of diphtheria toxin bound to nicotinamide adenine dinucleotide. Biochemistry. 35(4): 1137-1149, 1996. 79

[363] Ritter H., Koch-Nolte F., Marquez V.E. and Schulz G.E. Substrate binding and catalysis of ecto-ADP-ribosyltransferase 2.2 from rat. Biochemistry. 42(34):10155-10162, 2003. 79

[364] O'Neal C.J., Jobling M.G., Holmes R.K. and Hol W.G. Structural basis for the activation of cholera toxin by human ARF6-GTP. Science. 309(5737):1093-1096, 2005. 79

[365] Vogelsgesang M., Stieglitz B., Herrmann C., Pautsch A. and Aktories K. Crystal structure of the Clostridium limosum C3 exoenzyme. FEBS Lett. 582(7): 1032-1036, 2008. 79

[366] Evans H.R., Sutton J.M., Holloway D.E., Ayriss J., Shone C.C. and Acharya K.R. The crystal structure of C3stau2 from Staphylococcus aureus and its complex with NAD. J Biol Chem. 278(46): 45924-45930, 2003. 79

[367] Han S., Craig J.A., Putnam C.D., Carozzi N.B. and Tainer J.A. Evolution and mechanism from structures of an ADP-ribosylating toxin and NAD complex. Nat Struct Biol. 6(10): 932-936, 1999. 79

[368] Sundriyal A., Roberts A.K., Shone C.C. and Acharya K.R. Structural basis for substrate recognition in the enzymatic component of ADP-ribosyltransferase toxin CDTa from Clostridium difficile. J Biol Chem. 284(42): 28713-28719, 2009. 79

[369] Di Paola S., Micaroni M., Di Tullio G., Buccione R. and Di Girolamo M. PARP16/PARP15 Is a Novel Endoplasmic-Reticulum-Associated Mono-ADP-Ribosyltransferase That Interacts with, and Modifies Karyopherin-1. PLoS One. 7(6): e37352, 2012. 81

[370] Caldes C, Vilanova B, Adrover M, Munoz F, Donoso J. Understanding non-enzymatic aminophospholipid glycation and its inhibition. Polar head features affect the kinetics of Schiff base formation. Bioorg Med Chem. 19(15): 4536-4543, 2011. 82

[371] Fedorova M., Frolov A. and Hoffmann R. Fragmentation behavior of Amadori-peptides obtained by non-enzymatic glycosylation of lysine residues with ADP-ribose in tandem mass spectrometry. J Mass Spectrom. 45(6): 664-669, 2010. 82

[372] Janin J. Blind predictions of protein interfaces by docking calculations in CAPRI. Mol Biosyst. 6(12): 2351-2362, 2010. 82

[373] Scheeff E.D. and Bourne P.E. Structural evolution of the protein kinase-like superfamily. PLoS Comput Biol. 1(5): e49, 2005. 82

[374] Bothwell M. Functional interactions of neurotrophins and neurotrophin receptors. Annu. Rev. Neurosci. 18: 223-253, 1995. 83, 97

[375] Klein R., Jing S.Q., Nanduri V., O'Rourke E. and Barbacid M. The trk proto-oncogene encodes a receptor for nerve growth factor. Cell 65: 189-197, 1991. 83

[376] Capsoni S., Tiveron C., Vignone D., Amato G. and Cattaneo A. Dissecting the involvement of tropomyosin-related kinase A and p75 neurotrophin receptor signaling in NGF deficit-induced neurodegeneration. Proc. Natl. Acad. Sci. U.S.A. 107: 12299-12304, 2010 83

[377] Marchetti C. and Marie H. Hippocampal synaptic plasticity in Alzheimer's disease: what have we learned so far from transgenic models? Rev. Neurosci. 22: 373-402, 2011. 83

[378] Huang E.J. and Reichardt L.F. Trk receptors: Roles in neuronal signal transduction. Ann. Rev. Biochem. 72: 609-642, 2003. 83, 93

[379] Barker P.A. High affinity not in the vicinity? Neuron 53: 1-4, 2007. 83, 93

[380] Matrone C., Marolda R., Ciafre S., Ciotti M.T., Mercanti D., et al. Tyrosine kinase nerve growth factor receptor switches from prosurvival to proapoptotic activity via Abeta-mediated phosphorylation. Proc Natl Acad Sci U S A 106: 11358-11363, 2009. 83

[381] Foehr E.D., Lin X., O'Mahony A., Geleziunas R., Bradshaw R.A., et al. NF-kappa B signaling promotes both cell survival and neurite process formation in nerve growth factor-stimulated PC12 cells. J. Neurosci. 20: 7556-7563, 2000. 83

[382] Saragovi H.U., Zheng W.H., Maliartchouk S., DiGugliemo G.M., Mawal Y.R., et al. A TrkA-selective, fast internalizing nerve growth factor-antibody complex induces trophic but not neuritogenic signals. J. Biol. Chem. 273: 34933-34940, 1998. 83

[383] Koshland D.E., Nemethy G. and Filmer D. Comparison of Experimental Binding Data and Theoretical Models in Proteins Containing Subunits. Biochemistry 5: 365, 1966. 84

[384] Tsai C.J., Ma B.Y. and Nussinov R. Folding and binding cascades: Shifts in energy landscapes. Proc Natl Acad Sci U S A 96: 9970-9972, 1999. 84

[385] Kuzu G., Keskin O., Gursoy A. and Nussinov R. Expanding the conformational selection paradigm in protein-ligand docking. Methods Mol Biol 819: 59-74, 2012. 84

[386] Boehr D.D., Nussinov R. and Wright P.E. The role of dynamic conformational ensembles in biomolecular recognition. Nat. Chem. Biol. 5: 789-796, 2009. 84

[387] Bucher D., Grant B.J. and McCammon J.A. Induced fit or conformational selection? The role of the semi-closed state in the maltose binding protein. Biochemistry 50: 10530-10539, 2011. 84

[388] Nussinov R. and Ma B. Protein dynamics and conformational selection in bidirectional signal transduction. BMC Biol 10: 2, 2012. 84

[389] Guvench O., Hatcher E., Venable R.M., Pastor R.W. and MacKerell A.D. CHARMM Additive All-Atom Force Field for Glycosidic Linkages between Hexopyranoses. J. Chem. Theory Comput. 5: 2353-2370, 2009 85

[390] Essmann U., Perera L., Berkowitz M.L., Darden T., Lee H., et al. A Smooth Particle Mesh Ewald Method. J. Chem. Phys. 103: 8577-8593, 1995 85

[391] Miyamoto S. and Kollman P.A. Settle - an Analytical Version of the Shake and Rattle Algorithm for Rigid Water Models. J Comput Chem 13: 952-962, 1992 85

[392] Bonomi M., Branduardi D., Bussi G., Camilloni C., Provasi D., et al. PLUMED: A portable plugin for free-energy calculations with molecular dynamics. Comp. Phys. Commun. 180: 1961-1972, 2009 86

[393] Raiteri P., Laio A., Gervasio F.L., Micheletti C. and Parrinello M. Efficient reconstruction of complex free energy landscapes by multiple walkers metadynamics. J. Phys. Chem. B 110: 3533-3539, 2006. 86

[394] Bulheller B.M. and Hirst J.D. DichroCalc–circular and linear dichroism online. Bioinformatics 25: 539-540, 2009. 90

[395] Woody A.Y. and Woody R.W. Individual tyrosine side-chain contributions to circular dichroism of ribonuclease. Biopolymers 72: 500-513, 2003. 90

[396] Woo S.B., Whalen C. and Neet K.E. Characterization of the recombinant extracellular domain of the neurotrophin receptor TrkA and its interaction with nerve growth factor (NGF). Protein Sci 7: 1006-1016, 1998. 90, 97

[397] Dawbarn D., Fahey M., Watson J., Tyler S., Shoemark D., Sessions R., Zhang R., Brady L., Willis C. and Allen S.J NGF receptor TrkAd5: therapeutic agent and drug design target. 34(Pt 4): 587-590, 2006. 91

[398] Allen S.J., Watson J.J. and Dawbarn D. The neurotrophins and their role in Alzheimer's disease. Curr Neuropharmacol 9(4): 559-573, 2011. 91

[399] Ibanez C.F., Ebendal T., Barbany G., Murray-Rust J., Blundell T.L., et al. Disruption of the low affinity receptor-binding site in NGF allows neuronal survival and differentiation by binding to the trk gene product. Cell 69: 329-341, 1992. 91, 97

[400] Ibanez C.F., Ebendal T. and Persson H. Chimeric molecules with multiple neurotrophic activities reveal structural elements determining the specificities of NGF and BDNF. EMBO J 10: 2105-2110, 1991. 91, 97

[401] Kullander K. and Ebendal T. Neurotrophin-3 acquires NGF-like activity after exchange to five NGF amino acid residues: molecular analysis of the sites in NGF mediating the specific interaction with the NGF high affinity receptor. J Neurosci Res 39: 195-210, 1994. 91, 97

[402] Bibel M., Hoppe E. and Barde Y.A. Biochemical and functional interactions between the neurotrophin receptors trk and p75(NTR). Embo Journal 18: 616-622, 1999. 93

[403] Clary D.O. and Reichardt L.F. An Alternatively Spliced Form of the Nerve Growth-Factor Receptor Trka Confers an Enhanced Response to Neurotrophin-3. Proc Natl Acad Sci U S A 91: 11133-11137, 1994. 93

[404] Prieto C. and De Las Rivas J. APID: Agile Protein Interaction DataAnalyzer. Nucleic Acids Res 34: W298-302, 2006. 93

[405] Bourgeas R., Basse M.J., Morelli X. and Roche P. Atomic analysis of protein-protein interfaces with known inhibitors: the 2P2I database.PLoS One. 5(3): e9598, 2010. 95

[406] Thompson A.D., Dugan A., Gestwicki J.E. and Mapp A.K. Fine-Tuning Multiprotein Complexes Using Small Molecules. ACS Chem. Biol. 2012 [Epub ahead of print] 95

[407] Lorberboum-Galski H. Human toxin-based recombinant immunotoxins/chimeric proteins as a drug delivery system for targeted treatment of human diseases. Expert Opin Drug Deliv. 8(5):605-621, 2011. 95

[408] Allen S.J. and Dawbarn D. Clinical relevance of the neurotrophins and their receptors. Clin Sci (Lond). 110(2): 175-191, 2006. 97

[409] Counts S.E. and Mufson E.J. The role of nerve growth factor receptors in cholinergic basal forebrain degeneration in prodromal Alzheimer disease. J. Neuropathol. Exp. Neurol. 64(4): 263-272, 2005. 97

[410] He X. and Garcia K.C. Structure of nerve growth factor complexed with the shared neurotrophin receptor p75. Science 304: 870, 2004. 105

[411] Gong Y., Cao P., Yu H. and Jiang T. Crystal structure of the neurotrophin-3 and p75NTR symmetrical complex. Nature 454: 789-793, 2008. 105

[412] Radeke M.J., Misko T.P., Hsu C., Herzenberg L.A. and Shooter E.M. Gene transfer and molecular cloning of the rat NGF receptor. Nature 325: 593-597, 1987. 105

[413] Johnson D., Lanahan A., Buck C. R., Sehgal A., Morgan C. and Mercer E. Expression and structure of the human NGF receptor. Cell. 47: 545-554, 1986. 105

[414] Esposito D., Patel P., Stephens R.M., Perez P., Chao M.V., Kaplan D.R. and Hempstead B.L. The cytoplasmic and transmembrane domains of the p75 and Trk A receptors regulate high affinity binding to nerve growth factor. J Biol Chem 276: 32687-32695, 2001. 105

[415] Segal R.A. Selectivity in neurotrophin signalling: theme and variations. Ann. Rev. Neurosc. 26: 299-330, 2003. 105

[416] Ashkenazi A. Targeting death and decoy receptors of the tumour-necrosis factor superfamily. Nat. Rev. Cancer 2: 420-430, 2002. 105

[417] Frade J.M., Rodriguez-Tebar A. and Barde Y.A. Induction of cell death by endogenous nerve growth factor through its p75 receptor. Nature 383: 166-168, 1996. 105

[418] Friedman W.J. Neurotrophins induce death of hippocampal neurons via the p75 receptor. J Neurosci 20: 6340-6346, 2000. 105

[419] Brann A.B., Tcherpakov M., Williams I.M., Futerman A.H. and Fainzilber M. Nerve growth factor-induced p75-mediated death of cultured hippocampal neurons is age-dependent and transduced through ceramide generated by neutral sphingomyelinase. J Biol Chem 277: 9812-9818, 2002. 105

[420] Lee K.F., Davies A.M. and Jaenisch R. p75-deficient embryonic dorsal root sensory and neonatal sympathetic neurons display a decreased sensitivity to NGF. Development 120: 1027-1033, 1994. 105

[421] Bamji S.X., Majdan M., Pozniak C.D., Belliveau D.J., Aloyz R., Kohn J., Causing C.G. and Miller F.D. The p75 neurotrophin receptor mediates neuronal apoptosis and is essential for naturally occurring sympathetic neuron death. J Cell Biol 140: 911-923, 1998.. 105

[422] Troy C.M., Friedman J.E. and Friedman W.J. Mechanisms of p75-mediated death of hippocampal neurons: role of caspases. J Biol Chem 277: 34295-34302, 2002. 105

[423] Roux P.P. and Barker P.A. Neurotrophin signaling through the p75 neurotrophin receptor. Prog. Neurobiol. 67: 203-233, 2002. 93, 106

[424] Teng K.K. and Hempstead B.L. Neurotrophins and their receptors: signaling trios in complex biological systems. Cell Mol Life Sci. 61: 35-48, 2004. 106

[425] Crowder R.J. and Freeman R.S. Phosphatidylinositol 3-kinase and Akt protein kinase are necessary and sufficient for the survival of nerve growth factor-dependent sympathetic neurons. J Neurosci. 18(8): 2933-2943, 1998. 106

[426] Xing J., Kornhauser J.M., Xia Z., Thiele E.A. and Greenberg M.E. Nerve growth factor activates extracellular signal-regulated kinase and p38 mitogen-activated protein kinase pathways to stimulate CREB serine 133 phosphorylation. Mol Cell Biol. 18(4): 1946-1955, 1998. 106

Declaration

I herewith declare that I have produced this paper without the prohibited assistance of third parties and without making use of aids other than those specified; notions taken over directly or indirectly from other sources have been identified as such. This thesis has not previously been presented in identical or similar form to any other German or foreign examination board.

Curriculum vitae

Personal details

Date of birth: January 19th, 1982

Place of birth: Teramo, Italy

Nationality: Italian

Education

2009-2012: Doctorate at RWTH Aachen University, Germany

2007-2011: PhD in Structural and Functional Genomics, International School for Advanced Studies, SISSA/ISAS, Trieste, Italy

2003-2006: Master degree in Pharmaceutical Biotechnology, University of Modena and Reggio Emilia, Italy

2000–2003: Bachelor degree in Biotechnology, University of Modena and Reggio Emilia, Italy

1995-2000: High school diploma at the Liceo Classico Melchiorre Delfico, Teramo

Publications

1. Ferrari S., **Losasso V.**, Puneet S. and Costi M.P. Targeting the trypanosomatidic enzymes pteridine reductase and dihydrofolate reductase. In: Drug Discovery for Trypanosomatid Diseases .Edited by: Wiley-VCH Verlag GmbH and Co. KgaA Weinhein (DEU); T. Jager, O. Koch, L.Flohe. 2012; The RSC Drug Discovery Series; ISBN:9783527332557

2. **Losasso V.**, Schiffer S, Barth S. and Carloni P. Design of human Granzyme B variants resistant to Serpin B9. Proteins, doi: 10.1002/prot.24133, 2012.

3. **Losasso V.**, Pietropaolo A., Zannoni C., Gustincich S. and Carloni P. Structural role of compensatory aminoacid replacements in the alpha-synuclein protein. Biochemistry 50(32): 6994-7001, 2011.

4. Musiani F., Bertosa B., Magistrato A., Zambelli B., Turano P., **Losasso V.**, Micheletti C., Ciurli S. and Carloni P. Computational Study of the

DNA-Binding Protein Helicobacter pylori NikR: The Role of Ni2+ . J. Chem. Theory Comput., 6 (11): 3503-3515, 2010.

5. Latawiec D., Herrera F., Bek A., **Losasso V.**, Candotti M., Benetti F., Carlino E., Kranjc A., Lazzarino M., Gustincich S., Carloni P. and Legname G. Modulation of alpha-synuclein by dopamine analogs. PLoS One. 5(2): e2934, 2009.

6. Ferrari S., **Losasso V.** and Costi M.P. Sequence-based identification of specific drug target regions in the thymidylate synthase enzyme family. CHEMMEDCHEM, ISSN: 18607117, 2009.