
***In silico* characterisation of AtPARP1
and virtual screening for AtPARP inhibitors
to increase resistance to abiotic stress**

Dissertation

zur Erlangung des

Doktorgrades der Naturwissenschaften (Dr. rer. nat.)

der

Naturwissenschaftlichen Fakultät II – Chemie, Physik und Mathematik

der Martin-Luther-Universität

Halle-Wittenberg

vorgelegt von

Herr Diplom-Bioinformatiker

Peter Paul Heym

geboren am 22.08.1983 in Hoyerswerda

Die vorliegende Arbeit wurde im Zeitraum zwischen Februar 2009 und Dezember 2013 am Leibniz-Institut für Pflanzenbiochemie in Halle (Saale), in der Abteilung Natur- und Wirkstoffchemie unter der Betreuung von Prof. Dr. Ludger A. Wessjohann angefertigt.

1. Gutachter: Professor Dr. Ludger A. Wessjohann

2. Gutachter: Privatdozent Dr. Thomas E. Exner

Datum der Verteidigung: 12.12.2016

Edited by

This work was edited by:

Prof. Dr. L. A. Wessjohann (supervisor)

PD Dr. W. Brandt (group leader)

"In God we trust; all others must bring data."

William Edwards Deming (14.10.1900 – 20.12.1993)

Danksagung

Mein besonderer Dank gilt Professor Dr. L. A. Wessjohann für das in mich gesetzte Vertrauen, sowie die Diskussionsbereitschaft und das Einnehmen neuer Standpunkte bei scheinbar schwer-lösbaren Problemen. Für die Übernahme des Zweitgutachtens möchte ich mich bei Privatdozent Dr. Thomas E. Exner in aller Form bedanken.

Für die Übernahme des Mentoriats, als auch für fruchtbare Diskussionen zwischen den einzelnen Projektgruppen möchte ich mich bei Professor Dr. E. Peiter bedanken.

Bei PD Dr. Wolfgang Brandt möchte ich mich für seine ständige Diskussionsbereitschaft und seine Hilfe, nicht nur in wissenschaftlichen Fragen, bedanken. Ich danke ihm, wie auch Professor Wessjohann dafür, dass beide es mir ermöglicht haben, während meiner Promotion Beruf und Familie unter einen Hut zu bringen. Nicht zuletzt danke ich PD Dr. Brandt auch für seine Geduld, die es mir erlaubte, dieser Forschungsarbeit ihren eigenen Charakter zu verleihen.

Für die stetige Zusammenarbeit im Verbundprojekt möchte ich mich auch bei Dr. Torsten Geißler, Dagmar Rissel, Wiebke Zschesche, Dr. Heike Hahn, Professor Dr. K. Humbeck sowie Professor Dr. H.-J. Niclas bedanken.

Ganz besonders möchte ich mich bei Dr. Silke Pienkny bedanken, der die Proteinisolierung und Proteinreinigung gelang. Erst durch Ihre molekularbiologischen Arbeiten war es möglich, computerbasierte Ergebnisse zusammen mit *in vitro* Ergebnissen zu untersuchen. In diesem Zusammenhang danke ich auch Dr. Torsten Geißler, der sowohl den Enzym-Assay weiterführte, als auch den *Lemna*-Assay etablierte. Auch erst durch seine Arbeit lassen sich in dieser Arbeit vorgestellte Ergebnisse auf molekularbiologische Daten stützen.

Bei Dr. Robert Klein und Dr. Diana Schulze, deren abschließende Arbeiten ich zu Beginn meiner Promotion miterleben durfte, möchte ich danken für erste Einblicke, sowohl in die Promotion selbst, als auch den Stellenwert der Zeit des Zusammenschreibens. Bei meinen Kollegen Thomas Herberg und Eva Schulze möchte ich mich bedanken für die schöne Zeit im gemeinsamen Büro, in der es immer viel zu lachen gab, auch wenn es einem manchmal schwer fiel.

Weiterhin möchte ich mich bei meinen Kollegen Dr. Frank Broda, Dr. Susanne Aust, Juliane Fischer, Felix Rausch, Daniela Eisenschmidt, Anne Finck, Jennifer Szczesny, Richard Bartelt, Jördis-Ann Schüler, Michael Dressel und Susann Gruner bedanken für die schöne Zeit im Aquarium, für den tollen Zusammenhalt innerhalb der Gruppe, und den regen Wissensaustausch, durch den es immer mehr Lösungen als Probleme gab.

Bei Robert Berger möchte ich mich sowohl für seine Bereitschaft, neues zu entdecken, als auch die Wiederaufnahme des *Lemna*-Assays, bedanken, was jeweils dazu geführt hat, dass ich teilweise mehr Zeit im Labor, als am Computer verbrachte.

Weiterhin möchte ich mich bei allen Kollegen der Abteilung Natur- und Wirkstoffchemie des IPB Halle bedanken, in der jeder einzelne Mitarbeiter einen kleinen Teil dazu beigetragen hat, schöne und persönliche Erinnerungen an die Zeit im IPB mitzunehmen.

Auf Seiten des IPB möchte ich mich bei Frau Ines Stein für die unkomplizierte und freundliche Art und Weise bedanken, mit der sie organisatorische Angelegenheiten zu meistern vermag. Bei Jana Bähr und Heidrun Schweizer möchte ich mich für die problemlose Kommunikation mit dem AIP bedanken.

Ganz besonders möchte ich mich bei den Stickstoffwerken Piesteritz und dem Land Sachsen-Anhalt für die Finanzierung, als auch für den Erwerb der Testsubstanzen bedanken. Ohne diese wäre ein erheblicher Teil der Arbeit nicht möglich gewesen. Ganz speziell möchte ich mich in diesem Zusammenhang bei Prof. H.-J. Niclas bedanken, welcher mir zusammen mit PD Dr. Brandt und Prof. Wessjohann mehrfach die Chance ermöglichte, an hochklassigen Konferenzen teilzunehmen, die mich sowohl wissenschaftlich, in meiner beruflichen Laufbahn als auch im privaten Leben maßgeblich geprägt haben.

Bei meinen Eltern und „Schwiegereltern“ und meiner Familie möchte ich mich für die Unterstützung in allen Belangen bedanken.

Für die nahezu endlose Geduld, Ausdauer und Motivation, mich bei der Fertigstellung der Arbeit zu unterstützen, möchte ich mich bei meiner Partnerin, Andrea Brock, bedanken.

Bei meinen beiden Kindern möchte ich mich für ihre Art, mir zu zeigen, welche Sachen im Leben Priorität haben, bedanken. Bei beiden, als auch bei meiner Partnerin, möchte ich mich für das im Auge behalten des großen Zieles bedanken.

Abstract

In this work, a virtual screening (VS) workflow was developed for identification of compounds that lead to an increased drought stress resistance (DSR). The only verified targets that link DSR so far belong to the family of Poly (ADP-ribose) polymerase (PARP) enzymes. The inhibition of plant PARP is supposed to delay the breakdown of energy homeostasis during abiotic stress conditions. Therefore, a VS strategy to effectively screen commercial databases for plant PARP inhibitors was established. Inhibitory effects of VS-proposed compounds were tested on purified *Arabidopsis thaliana* L. PARP1 protein (*At*PARP1) *in vitro* and on *Lolium perenne* plants as monocotyledons to verify the hypothesis.

The developed VS strategy was based on human PARP1 (*Hs*PARP1) which is a key target in (e.g. ovarian and breast) cancer therapy. For *Hs*PARP1, several crystal structures and a wide knowledge of inhibitors are available. Based on *Hs*PARP1, homology models of *At*PARP1 and *At*PARP2 were derived. Subsequently, the knowledge about known human PARP1 inhibitors and decoys was applied to statistical methods including receiver-operating characteristics and statistical power analysis. Extensive docking studies with statistical methods were conducted to define a docking score threshold to effectively discriminate potential inhibitors from decoy structures. The threshold was subsequently adjusted for *At*PARP1, again using statistical hypotheses and methods of inference. These methods allowed for prediction of the performance of the VS route on a commercial database which was screened for *At*PARP inhibitors. The number of resulting structures was reduced by applying the docking score threshold. Eventually, 121 compounds were selected and tested *in vitro* on *At*PARP1. Among those, 47 compounds were found to be inhibiting *At*PARP1, corresponding to a hit rate of about 39%. Out of these 47 compounds, 33 were predicted to be inhibiting according to the docking score threshold.

Furthermore, for 52 of the tested compounds, the IC_{50} values were determined. Among those, 6 compounds showed an IC_{50} below 1 μ M, 26 compounds exhibited an IC_{50} of less than 10 μ M. Among 22 compounds which were tested *in planta* on *Lolium perenne* plants, 9 showed a positive effect on dry mass production under drought stress.

Apart from the VS for *At*PARP1 inhibitors the catalytic domains of *At*PARP1 and *At*PARP2 were characterised *in silico*. The characterisation comprised analysis of protein

quality as a result of the homology modelling process. Protein stability was investigated by comparing molecular dynamics (MD) simulation data with experimentally determined data from other PARP orthologues. Multiple-step homology modelling together with MD simulation were used to investigate the natural substrate binding of *At*PARP1. Based on the *in silico* characterisation of *At*PARP the VS could be performed. Finally, experimentally determined IC₅₀ values for VS-proposed *At*PARP1 inhibitors and molecular descriptors were used to derive binary quantitative structure-activity relationships (binary QSAR).

The research shows that PARP1 is involved in the regulation of abiotic stress response in *Arabidopsis thaliana*. I developed a virtual screening route for *At*PARP1 based on the knowledge about human PARP by applying statistical methods. Although docking protocols are thought to be unable to predict the activity of compounds from the docking score, I showed that at least an effective discrimination of inhibitors from non-binders can be possible, if statistical assumptions are taken into account.

Zusammenfassung

In dieser Arbeit wurde eine virtuelle Screening (VS) Prozedur entwickelt, die der Identifizierung von Verbindungen dient, welche die Toleranz gegenüber Trockenstress bei Pflanzen erhöhen sollte. Die bislang einzig verifizierten Pflanzenproteine, die in Verbindung mit einer erhöhten Trockenstresstoleranz stehen, gehören allesamt der Familie der Poly-(ADP-ribose)-polymerasen (PARP) an. Dabei wird vermutet, dass eine Inhibierung von PARP Proteinen während abiotischer Stressbedingungen zu einem verzögerten Zusammenbruch der Energiehomöostase der Pflanze führt. Ziel dieser Arbeit war es daher, eine VS Strategie zu entwickeln, welche es erlaubt, kommerzielle Datenbanken effektiv nach potentiellen pflanzlichen PARP Inhibitoren zu durchsuchen. Effekte, die durch potentielle PARP Inhibierung hervorgerufen werden könnten, wurden an gereinigtem *Arabidopsis thaliana* L. PARP1 protein (*AtPARP1*) *in vitro* sowie an *Lolium perenne* Pflanzen als Vertreter der Monokotyledonen getestet, um die Hypothesen zu testen.

Die hier entwickelte VS Strategie nutzt das Wissen über humane PARP1 (*HsPARP1*) Inhibitoren, da dieses Protein ein potentielles *target* in der Krebsbekämpfung (u.a. von Ovarialkarzinomen und Brustkrebs) darstellt. Für *HsPARP1* existieren bereits mehrere Röntgenkristallstrukturen, sowie breites Wissen über *HsPARP1* Inhibitoren und zu diesen strukturell verwandte Verbindungen, die jedoch nicht an *HsPARP1* binden (sogenannte *decoys*). Basierend auf den Röntgenkristallstrukturen von *HsPARP1* wurden Homologiemodelle von *AtPARP1* und *AtPARP2* erstellt. Darüber hinaus wurden im Rahmen von umfangreichen Docking-Analysen von *HsPARP1* Bindern und Nicht-Bindern statistische Verfahren wie *receiver operating characteristics* und Power Analysen angewendet. Diese erlaubten eine effektive Unterscheidung tatsächlicher Inhibitoren von *decoys* unter Einbeziehung von *docking score* Grenzen. Während diese Grenzen vom humanen PARP1 resultierten, wurde diese unter Einhaltung entsprechender statistischer und biologischer Annahmen auf *AtPARP1* und *AtPARP2* angepasst und übertragen. Diese Grenze sollte eine effektive Suche in kommerziellen Datenbanken nach potentiellen *AtPARP* Inhibitoren ermöglichen.

Zusammen mit einem ebenfalls auf *HsPARP1* basierenden Pharmakophor wurde anschließend eine Datenbank mit mehr als 40.000 Strukturen durchsucht und an Hand der Suchkriterien 121 Verbindungen ausgewählt und am *AtPARP1* Enzym *in vitro* getestet

wurden. Von diesen waren insgesamt 47 *At*PARP1-aktiv, was einer Erfolgsrate von rund 39% entspricht. Von diesen 47 Aktiven wurden 33 anhand der festgelegten *docking score* grenze als aktiv vorhergesagt. Weiterhin konnten von 52 der 121 Verbindungen die IC₅₀ Werte ermittelt werden. Von den untersuchten Verbindungen wiesen 6 einen Wert im nanomolaren Bereich, weitere 20 Werte unter 10 µM auf. Von insgesamt 22 getesteten Verbindungen an *Lolium perenne* Pflanzen zeigten 9 im Vergleich zu Kontrollpflanzen einen positiven Effekt unter Trockenstressbedingungen.

Neben dem virtuellen Screening wurden die dafür verwendeten Homologiemodelle der katalytischen Domänen vom *At*PARP1 und *At*PARP2 auf deren *in silico*-Qualität hin untersucht. Die Qualität der Modelle wurde dabei verglichen mit der Qualität von Röntgenkristallstrukturen von PARP-Orthologen der Proteindatenbank (PDB), wobei der Einfluss von Inhibitoren auf die Proteinstabilität, sowie der Bindemodus der natürlichen Substrate von PARPs untersucht worden. Diese Untersuchungen wurden mit Hilfe von Moleküldynamik (MD)-Simulationen durchgeführt und ebenfalls statistisch ausgewertet. Dabei zeigte sich, dass die Ergebnisse, die durch Röntgenkristallstrukturen von PARP-Orthologen festgestellt worden sind, in analoger Weise bei Homologiemodellen von *At*PARP1 ebenfalls beobachtet werden können. Damit wurden weitere Indizien gefunden, die nahelegen, dass die Qualität der Homologiemodelle und der verwendeten Screening Methoden ausreichend sind, um effektiv nach neuen Inhibitoren suchen zu können.

Abschließend wurde mit Hilfe von binären quantitativen Struktur-Wirkungs-Beziehungen (*binary QSAR*) untersucht, welche Eigenschaften (beschrieben durch molekulare Deskriptoren) der als aktiv und nicht aktive getesteten Inhibitoren für deren Aktivität bzw. Nicht-Aktivität verantwortlich sind. Das resultierende binäre QSAR Modell zeigte eine hohe Sensitivität und Spezifität und kann damit zum weiteren Verständnis der Bindung von Strukturen an *At*PARP1 *in silico* beitragen.

Die Ergebnisse der vorgelegten Arbeit zeigen, dass *Arabidopsis thaliana* PARP1 in die Regulation der abiotischen Stressantwort involviert ist. Sie legt dar, dass molekulare Modellierungs-Studien die experimentellen Ergebnisse der *in vitro* und *in vivo* Studien zu pflanzlichen PARP Inhibitoren unterstützen und erklären können. Weiterhin wird gezeigt, dass die Proteinmodelle von *At*PARP ähnliche Qualität aufweisen wie orthologe Röntgenkristallstrukturen und damit ähnliche Erkenntnisgewinne durch molekulare Modellierungs-Studien möglich sind wie bei Röntgenstrukturen. Sie ist die erste Arbeit, in der mit Hilfe des virtuellen Screenings neue Inhibitoren für *At*PARP gefunden wurden.

Table of contents

Edited by	3
Danksagung	5
Abstract	7
Zusammenfassung	9
Table of Abbreviations	17
List of colors	19
List of Tables	20
Table of Figures	21
1 Introduction	23
1.1 The family of Poly (ADP-ribose) polymerases (PARP)	23
1.1.1 Human PARP	23
1.1.2 Plant PARP	24
1.1.3 The catalytic reaction of PARP	25
1.2 <i>Hs</i> PARP1 as a therapeutic target	28
1.2.1 <i>Hs</i> PARP1 and DNA repair	28
1.2.2 <i>Hs</i> PARP1 and synthetic lethality	29
1.2.3 <i>Hs</i> PARP1 and ischemia	30
1.3 PARP inhibitors	31
1.3.1 Development of <i>Hs</i> PARP inhibitors	31
1.3.2 <i>Hs</i> PARP inhibitors in clinical trials	32
1.4 The role of poly (ADP-ribosyl)ation in plants	35
1.4.1 Plant PARP	35
1.4.2 Plant PARG	36
1.4.3 Plant NUDX	36
1.5 Virtual screening in Lead Discovery	38
1.5.1 Virtual screening for human PARP inhibitors	39
1.5.2 Virtual screening in agrochemistry	41
1.6 Aim of this work	43
2 Materials & Methods	45
2.1 Data sets	45
2.1.1 Natural substrate and natural substrate analogues	45
2.1.2 Commercial database	46
2.1.3 Known human PARP inhibitors	46
2.1.4 Known human PARP decoys	47

2.2	Sequence analyses.....	47
2.2.1	Pairwise sequence alignment.....	47
2.2.2	Multiple sequence alignment.....	47
2.3	Programs for Homology Modelling and Docking.....	48
2.3.1	Molecular Operating Environment (MOE).....	48
2.3.2	POSIT.....	48
2.3.3	YASARA.....	49
2.3.3.1	Homology modelling in YASARA.....	50
2.3.3.2	MD simulations in YASARA.....	50
2.3.4	ConfGen.....	51
2.3.5	LigPrep.....	51
2.3.6	Glide.....	52
2.3.7	GOLD.....	52
2.3.8	PLANTS.....	53
2.4	Application-dependent Homology Modelling.....	54
2.4.1	Investigation of protein stability.....	54
2.4.2	Investigation of positions of natural substrates.....	55
2.4.3	Investigation of bioactive conformation of <i>At</i> PARP inhibitors.....	57
2.5	Pharmacophore creation.....	59
2.6	Docking procedure.....	60
2.6.1	General aspects.....	60
2.6.2	Data sets.....	61
2.6.3	Docking programs.....	61
2.6.4	PARP pharmacophore-directed docking.....	63
2.6.5	Definition of a correct docking pose.....	64
2.7	Methods of probability and inference.....	65
2.7.1	Null hypothesis significance testing and statistical power.....	65
2.7.2	One-sample <i>t</i> -test.....	67
2.7.3	Unpaired two-sample <i>t</i> -test.....	67
2.7.4	Cumulative distribution function (cdf).....	67
2.7.5	Pearson's Chi-squared test.....	68
2.7.6	Binary quantitative structure-activity relationship.....	69
2.7.6.1	Binary quantitative structure-activity relationship terminology.....	69
2.7.6.2	Binary QSAR evaluation.....	70
2.7.7	Receiver Operator Characteristics.....	72
2.7.8	Analysis of MD simulations.....	73
3	Results.....	74
3.1	Sequence analysis.....	74
3.2	Homology Modelling.....	76
3.2.1	Evaluation of homology models.....	78
3.2.2	Model refinement.....	80
3.2.2.1	RAMPAGE.....	81

3.2.2.2	ProSA-web and Errat	81
3.2.2.3	Verify3D.....	82
3.3	Investigation of protein stability	83
3.3.1	Overall shape of B-factor and C _α distributions.....	83
3.3.2	Local structural shifts upon inhibitor binding.....	86
3.4	Investigation of positions of natural substrates.....	87
3.4.1	Homology model of natural substrate-bound <i>At</i> PARP1	87
3.4.2	Positioning of the nicotinamide moiety of NAD ⁺	90
3.4.3	Binding of the adenine moiety of the donor structure NAD ⁺	92
3.4.4	The role of the catalytic glutamate.....	93
3.4.5	Positioning of the adenine moiety of the acceptor structure.....	97
3.5	Docking.....	101
3.5.1	Docking program selection	101
3.5.2	Receiver Operator Characteristics (ROC) curve	102
3.5.3	Inference for data set docking score distributions	104
3.5.4	Normal approximation of docking score distributions	104
3.5.5	<i>Hs</i> PARP1 inhibitor docking score threshold derivation.....	106
3.5.6	Derivation of docking score threshold for <i>At</i> PARP.....	113
3.5.6.1	Assumptions.....	113
3.5.6.2	Differences of the docking procedure between <i>Hs</i> PARP1 and <i>At</i> PARP1	113
3.6	Processing data from a commercial data base	118
3.7	Pharmacophore filtering	119
3.7.1	Pharmacophore selectivity.....	119
3.7.2	Pharmacophore filtering of a commercial database.....	120
3.7.3	Pharmacophore filtering of structures with specific core structures.....	120
3.8	Selection of compounds.....	121
3.8.1	Compound selection based on docking score and pharmacophore selection	121
3.8.2	Compound selection based on chemical characteristics.....	121
3.8.3	Analysis of docking results / retrospective power analysis	125
3.9	Finding most probable pose of active compounds.....	129
3.10	Structure-activity relationship.....	136
3.10.1	General aspects	136
3.10.2	Binary structure-activity relationship	138
3.10.3	Influence of inhibitors on plant dry mass production	144
4	Summary and outlook	146
5	Appendix	148
5.1	Pairwise sequence alignments	148
5.1.1	Needle settings.....	148
5.1.2	<i>At</i> PARP1 – <i>At</i> PARP2	148
5.1.3	<i>At</i> PARP1 – <i>At</i> PARP3	149
5.1.4	<i>At</i> PARP1 – <i>Hs</i> PARP1	149

5.1.5	<i>At</i> PARP1 – <i>Hs</i> PARP2	150
5.1.6	<i>At</i> PARP1 – <i>Hs</i> PARP3	150
5.1.7	<i>At</i> PARP1 – <i>Gg</i> PARP1	151
5.1.8	<i>At</i> PARP2 – <i>At</i> PARP3	151
5.1.9	<i>At</i> PARP2 – <i>Hs</i> PARP1	152
5.1.10	<i>At</i> PARP2 – <i>Hs</i> PARP2	152
5.1.11	<i>At</i> PARP2 – <i>Hs</i> PARP3	153
5.1.12	<i>At</i> PARP2 – <i>Gg</i> PARP1	153
5.1.13	<i>At</i> PARP3 – <i>Hs</i> PARP1	154
5.1.14	<i>At</i> PARP3 – <i>Hs</i> PARP2	154
5.1.15	<i>At</i> PARP3 – <i>Hs</i> PARP3	155
5.1.16	<i>At</i> PARP3 – <i>Gg</i> PARP1	155
5.1.17	<i>Hs</i> PARP1 – <i>Hs</i> PARP2.....	156
5.1.18	<i>Hs</i> PARP1 – <i>Hs</i> PARP3.....	156
5.1.19	<i>Hs</i> PARP1 – <i>Gg</i> PARP1	157
5.1.20	<i>Hs</i> PARP2 – <i>Hs</i> PARP3.....	157
5.1.21	<i>Hs</i> PARP2 – <i>Gg</i> PARP1	158
5.2	Multiple Sequence Alignment	159
5.2.1	Clustal Omega settings	159
5.2.2	Multiple Sequence Alignment I.....	159
5.2.3	Multiple Sequence Alignment II.....	160
5.3	Pharmacophore annotations.....	161
5.4	Homology modelling	162
5.4.1	YASARA homology modelling report: <i>At</i> PARP1	162
5.4.1.1	Report section 1: Homology modelling target	162
5.4.1.2	Report section 4: Secondary structure prediction	162
5.4.1.3	Report section 5: Initial homology models	163
5.4.1.4	Report section 6 and 7: Model ranking and hybrid model.....	164
5.4.2	YASARA homology modelling report: <i>At</i> PARP2	165
5.4.2.1	Report section 1: Homology modelling target	165
5.4.2.2	Report section 4: Secondary structure prediction	165
5.4.2.3	Report section 5: Initial homology models	166
5.4.2.4	Report section 6s and 7: Model ranking and hybrid model	167
5.4.3	YASARA homology modelling report: <i>At</i> PARP1 (protein stability)	168
5.4.3.1	Report section 1: Homology modelling target	168
5.4.3.2	Report section 4: Secondary structure prediction	168
5.4.3.3	Report section 5: The target sequence profile (excerpt)	169
5.4.3.4	Report section 6: The initial homology models	171
5.4.3.5	Report sections 7 and 8: Model ranking and the hybrid model	178
5.4.4	YASARA homology modelling report: <i>At</i> PARP1 (natural substrates).....	179
5.4.4.1	Report section 1: Homology modelling target	179
5.4.4.2	Report section 4: Secondary structure prediction	179

5.4.4.3	Report section 5: The target sequence profile (excerpt)	180
5.4.4.4	Report section 6: The initial homology models	182
5.4.4.5	Report sections 7 and 8: Model ranking and the hybrid model	187
5.4.5	Homology model structure evaluation.....	188
5.5	Model refinement.....	190
5.5.1	ProSA-web and Errat	190
5.5.2	RAMPAGE.....	191
5.5.3	Verify 3D	192
5.6	Docking.....	193
5.6.1	Comparison of docking protocols.....	193
5.6.2	Docking score distributions	194
5.6.3	Inference of docking score distributions.....	194
5.6.4	Observed and approximated docking scores, differences.....	195
5.6.5	Number of structures found in pose 1.....	195
5.6.6	Number of structures docked correctly in docking runs.....	196
5.6.7	Pearson's Chi-squared test results	196
5.6.8	PLANTS script for docking.....	197
5.6.9	R script for ROC curves.....	198
5.6.10	Docking results – confidently docked inhibitors	199
5.7	Screening results	200
5.7.1	Pharmacophore screening results.....	200
5.8	ProBiS – conserved amino acids.....	200
5.9	Inhibitors.....	201
5.10	Inhibitors; <i>in silico</i> and <i>in vitro</i> results	219
5.10.1	Docking results, <i>in silico</i> and <i>in vitro</i> results.....	219
5.10.2	POSIT results	223
5.10.3	Structures used for binary QSAR – training set - actives	224
5.10.4	Structures used for binary QSAR – training set – inactives	224
5.10.5	Binary QSAR – model selection.....	226
5.10.6	Binary QSAR – training set– results.....	227
5.10.7	Binary QSAR – external validation – results.....	228
5.10.8	Binary QSAR PC analysis	228
5.11	MD simulation and analysis	229
5.11.1	YASARA script - converting YASARA snapshots to .PDB files.....	229
5.11.2	R script for preparing MD analysis.....	229
5.11.3	RMSD analysis of NAD-CNA-ligated <i>At</i> PARP1.....	230
5.12	<i>Lolium perenne</i> screening results	232
	References	233
	Curriculum vitae	245
	Publications.....	246

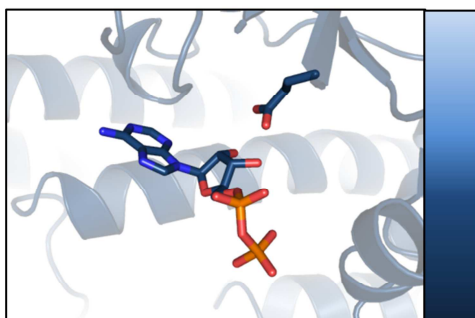
Eidesstattliche Erklärung	248
Statement under oath	248

Table of Abbreviations

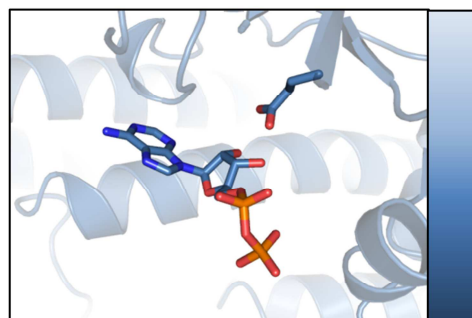
3MB	3-aminobenzamide
4AN	4-amino-1,8-naphthalimide
aa	amino acid
ABA	abscisic acid
AMD	automodification domain
ACO	ant colony optimisation
ADPR	ADP-ribose
ADPRT	ADP-ribosyl transferase
AHAS	acetoxy acid synthase
ALS	acetolactate synthase
APP	<i>Arabidopsis thaliana</i> homologue of PARP
(m)ART	(mono) ADP-ribosyltransferase
<i>At</i>	<i>Arabidopsis thaliana</i> L.
<i>BRCA1</i>	breast cancer 1, early onset gene
BRCA1	breast cancer 1, early onset gene product
BRCT	BRCA1 C-terminus
CD	catalytic domain
CDF	cumulative distribution function
CI	confidence interval
CNA	carba nicotinamide adenine dinucleotide
DBD	DNA-binding domain
DNA	deoxyribonucleic acid
DSB	double- strand break
EC	enzyme classification
EF	enrichment factor
FRQ	5-Fluoro-1-[4-(4-phenyl-3,6-dihydropyridin-1(2H)-yl)butyl]quinazoline-2,4(1H,3H)-dione
<i>Gg</i>	<i>Gallus gallus</i> (Chicken)
HR	homologous recombination
<i>Hs</i>	<i>Homo sapiens</i>
IC ₅₀	inhibitor concentration at which enzyme activity is reduced by 50%

K_i	inhibition constant
LOO	leave one out
MD	molecular dynamics
MLR	multiple linear regression
<i>Mm</i>	<i>Mus musculus</i> (Mouse)
MMR	mismatch repair
NA	nicotinamide
NAD ⁺	nicotinamide adenine dinucleotide
NER	nucleotide excision repair
NF- κ B	nuclear factor kappa-light-chain-enhancer of activated B cells
NHST	null hypothesis significance testing
NHEJ	nonhomologous end joining
PARP	Poly(ADP-ribose)Polymerase
PARG	Poly(ADP-ribose)Glycohydrolase
PC	principal component
PEF	pharmacophore enrichment factor
PCD	programmed cell death
PLANTS	Protein Ligand ANT System
PDB	Protein Data Bank
PPO	protoporphyrinogen oxidase
PSII	photosystem II
RI	reperfusion injury
RMSE	root mean squared error
ROC	receiver operator characteristics
ROS	reactive oxygen species
SAP	SAF-A/B, Acinus, and PIAS
SRS	simple random sample
SSB	single-strand break
SSBR	single-strand break repair
TMZ	Temozolomide
VS	virtual screening

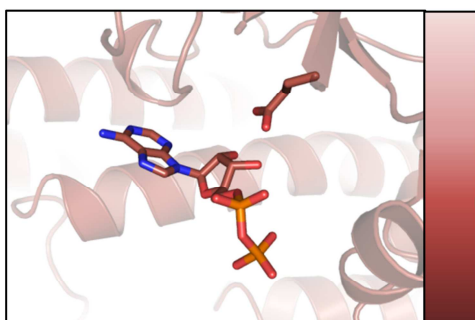
List of colors



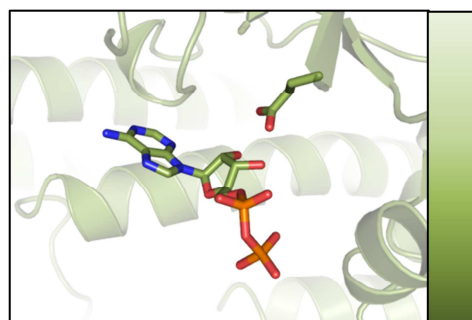
representation of *At*PARP1



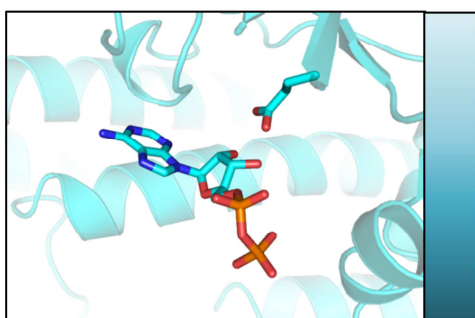
representation of *At*PARP2



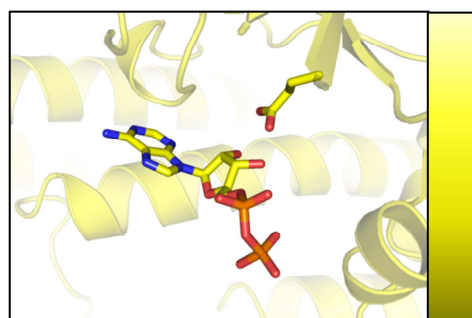
representation of *Hs*PARP1/2/3
especially *Hs*PARP1 ligands



representation of *Hs*PARP1
especially *Hs*PARP1 decoys



representation of other ADPRT
especially Diphtheria toxin



representation of *Gg*PARP1

List of Tables

Table 1.1: HsPARP1 inhibitors in clinical trials	34
Table 1.2: Classification of Herbicides according to HRAC	42
Table 2.1: Homology modelling steps in YASARA	50
Table 2.2: Steps to be performed to define an AtPARP docking procedure	60
Table 2.3: Steps necessary to perform null hypothesis significance testing (NHST)	65
Table 2.4: Relationship between Type I and Type II errors, and statistical power	66
Table 3.1: General information about catalytic domains (CD) in selected PARP	74
Table 3.2: Sequence similarities and sequence identities for selected PARP	74
Table 3.3: Evaluation of homology models	79
Table 3.4: Summary of AtPARP1 and AtPARP2 model refinement results	82
Table 3.5: Summary statistics for MD simulations	83
Table 3.6: ProBiS results	88
Table 3.7: Results of MD simulations - comparison with experimental data I	90
Table 3.8: Results of MD simulations - comparison of experimental data II	92
Table 3.9: Results of MD simulations - comparison of experimental data III	95
Table 3.10: Results of MD simulations - comparison of experimental data IV	98
Table 3.11: Characteristics of the docking threshold	107
Table 3.12: Compound selection based in different focal points	109
Table 3.13: Statistical vs. practical significance	110
Table 3.14: Characteristics of docking procedure	114
Table 3.15: AtPARP1 and AtPARP2 docking score differences	115
Table 3.16: Derivation of new docking thresholds for AtPARP1 and AtPARP2	117
Table 3.17: PARP Pharmacophore selectivity	119
Table 3.18: Compound selection: selected structures and their availability	121
Table 3.19: Structures selected for virtual screening: quinazolinones	122
Table 3.20: Structures selected for virtual screening: phthalazinones	123
Table 3.21: Structures selected for in vitro screening: other chemical classes	124
Table 3.22: Retrospective power analysis	125
Table 3.23: PARP crystal structures used with POSIT	129
Table 3.24: Results of POSIT application I	131
Table 3.25: Results of POSIT application II	132
Table 3.26: Selected descriptors for binary QSAR	139
Table 3.27: Results of binary QSAR	140

Table of Figures

Figure 1.1: Catalytic reaction of PARP	26
Figure 1.2: The proposed mechanism of the branching and elongation reaction of PARP	27
Figure 1.3: Synthetic lethality and PARP inhibition	30
Figure 1.4: Examples of developed PARP inhibitors.....	32
Figure 1.5: Examples of human PARP1 inhibitors having entered clinical trials.....	33
Figure 1.6: Interplay between PARP, PARG and NUDX proteins in abiotic stress	37
Figure 1.7: Workflow and aim of this work	44
Figure 2.1: Natural substrate of HsPARP1 and substrate analogue carba-NAD.....	45
Figure 2.2: Examples of HsPARP1 inhibitors.....	46
Figure 2.3: Examples of HsPARP1 decoys	47
Figure 2.4: POSIT probability map, modified from POSIT manual	49
Figure 2.5: Workflow of investigation of protein stability in AtPARP1.....	55
Figure 2.6: Workflow using ProBiS	56
Figure 2.7: Schematic representation of ProBiS algorithm I.....	56
Figure 2.8: Schematic representation of ProBiS algorithm II	57
Figure 2.9: Finding most probable bioactive conformations of AtPARP inhibitors.....	58
Figure 2.10: The PARP pharmacophore	59
Figure 2.11: Docking workflow for establishing an AtPARP docking procedure.....	61
Figure 2.12: Hydrogen bond weights adjusted for pharmacophore-directed docking	63
Figure 2.13: Definition of a correct docking pose.....	64
Figure 2.14: NHST and statistical power	66
Figure 2.15: Example of an ROC	72
Figure 3.1: Multiple sequence alignment between selected PARP	75
Figure 3.2: Protein models of HsPARP1 and AtPARP1/2.....	77
Figure 3.3: Results of ProSA-web.....	80
Figure 3.4: Analysis of MD simulations with unligated and 4AN-ligated AtPARP1 I	84
Figure 3.5: Analysis of MD simulations with unligated and 4AN-ligated AtPARP1 II.....	85
Figure 3.6: Analysis of MD simulations with unligated and 4AN-ligated AtPARP1 III....	86
Figure 3.7: ProBiS results I - Diphtheria Toxin and GgPARP1 active sites superposition	89
Figure 3.8: Positioning of the nicotinamide moiety of NAD ⁺ in AtPARP1.....	91
Figure 3.9: Positioning of the adenine moiety of NAD ⁺ in AtPARP1	93
Figure 3.10: Positioning of the nicotinamide moiety of NAD ⁺ in AtPARP1.....	94
Figure 3.11: The role of the catalytic glutamate in the catalytic reaction	96
Figure 3.12: Positioning adenine moiety of CNA in GgPARP1 and AtPARP1	97
Figure 3.13: AtPARP1 MD simulation analysis, adenine moiety of CNA positioning I....	98
Figure 3.14: AtPARP1 MD simulation analysis, adenine moiety of CNA positioning II ..	99

Figure 3.15: Comparison of docking programs.....	101
Figure 3.16: ROC curves of PLANTS docking protocols I and II.....	102
Figure 3.17: Docking score distributions – normal approximation I	104
Figure 3.18: Docking score distributions – normal approximation II.....	105
Figure 3.19: Normal approximation III	106
Figure 3.20: Number of structures being docked into HsPARP1 and AtPARP.....	114
Figure 3.21: HsPARP1 and HsPARP2 docking scores of 142 HsPARP1 inhibitors.....	116
Figure 3.22: Schematic representation of subsites S1 – S4 and their occupancies	122
Figure 3.23: Retrospective power analysis.....	126
Figure 3.24: Finding probable bioactive conformations of new AtPARP1 inhibitors	130
Figure 3.25: POSIT results: quality of prediction for AtPARP1 inhibitors.	132
Figure 3.26: POSIT results: quality of prediction for AtPARP1 inhibitors.	134
Figure 3.27: Binary QSAR workflow	138
Figure 3.28: Binary QSAR results I	141
Figure 3.29: Binary QSAR results II – External validation	142
Figure 5.1: RAMPAGE results of PDB entry 1UK1	188
Figure 5.2: RAMPAGE results of YASARA model of 1UK1.....	188
Figure 5.3: RAMPAGE results of homology model of AtPARP1.....	189
Figure 5.4: RAMPAGE results of homology model of AtPARP2.....	189
Figure 5.5: RAMPAGE - evaluation of AtPARP1 initial model with MD-refinement	191
Figure 5.6: RAMPAGE - evaluation of AtPARP2 initial model with MD-refinement	191
Figure 5.7: Verify 3D - AtPARP1 and AtPARP2 initial models and MD-refinement	192
Figure 5.8: Docking score distributions – normal approximation III.....	194
Figure 5.9: HsPARP1 and AtPARP2 docking scores of 142 HsPARP1 inhibitors	199
Figure 5.10: Substructure search for phalazinone and quinazolinone.....	200
Figure 5.11: Results of MD simulations of NAD-CNA-ligated AtPARP1.....	230
Figure 5.12: Results of MD simulations of NAD-CNA-ligated AtPARP1.....	230
Figure 5.13: Results of MD simulations of NAD-CNA-ligated AtPARP1.....	231
Figure 5.14: Results of MD simulations of NAD-CNA-ligated AtPARP1.....	231
Figure 5.15: Lolium perenne relative dry mass production for 22 AtPARP1 inhibitors...	232

1 Introduction

1.1 The family of Poly (ADP-ribose) polymerases (PARP)

1.1.1 Human PARP

Poly (ADP-ribose)-polymerases (PARP, EC 2.4.2.30), which are also called Diphtheria Toxin-like ADP-ribosyltransferases (ADRT), are nuclear and cytosolic enzymes that are mainly responsible for the synthesis of negatively charged poly(ADP-ribose) polymers. ADP-ribose moieties are formed by the cleavage of PARP's natural substrate β -nicotinamide adenine dinucleotide (NAD^+), in which nicotinamide (NA) is released as a reaction by-product (Figure 1.1). ADP-ribose monomers are covalently attached to target acceptor proteins and formation of further ADP-ribose units leads to the accumulation of poly(ADP-ribose) (PAR) polymers. The process of poly(ADP-ribosylation) is a post-translational modification and is involved in several biological processes that include DNA repair, cellular signaling, transcription, cell-cycle regulation, and mitosis. Hence, PARP plays an important role in inflammation, cancer, differentiation, stress response and development.

In 1956, it was discovered that DNA-alkylating agents caused depletion in the NAD^+ content in human ascites-tumour cells.¹ In the 1960s this observation was attributed to an enzyme which today is known as PARP.² PARP-like genes have been identified in all eukaryotes (except for *S. cerevisiae* and *S. pombe*), archaeobacteria, eubacteria and double-stranded DNA viruses. PARP enzymes constitute a superfamily, all containing a PARP catalytic site, that is denoted as the PARP signature.³⁻⁵ In the human genome, 17 members of PARP (*Hs*PARP) have been identified so far.⁶⁻⁹ PARP1-5 show catalytic activity and all contain a conserved glutamate residue responsible for catalytic activity. PARP 6-8, 10-12 and 14-16 are confirmed or putative mono(ADP-ribosyl) transferases (mARTs). PARP9 and PARP13 lack the catalytic glutamate and NAD^+ binding residues and are likely inactive.¹⁰ All PARP members consist of several independently folded domains.

By the whole of human PARP members, Poly (ADP-ribose) polymerase 1 (*Hs*PARP1) is investigated most rigorously. It is a protein of approximately 113 kDa¹¹ and it accounts for the about 90% of PAR production.¹² The domains of *Hs*PARP1 include an *N*-terminal DNA-binding domain (DBD), an automodification domain (AMD) and a *C*-terminal catalytic

domain (CD).^{7,9} The DBD contains three zinc fingers (Zn1/FI, Zn2/FII, Zn3/FIII) that mediate binding to DNA and interdomain contacts that are important for DNA-dependent enzyme activation.^{13,14} A nuclear localisation signal (NLS) and a caspase-3 cleavage site are localized at the DBD.^{7,9} The AMD acts as an acceptor of PAR during the automodification of PARP.¹⁵ A breast cancer 1, early onset gene product (BRCA 1) C-terminus (BRCT) fold is contained in the AMD, responsible for the mediation of protein-protein interactions with DNA repair enzymes. The most conserved domain across the PARP family is the CD. It contains the PARP signature and the active site where NAD⁺ is bound. The CD contains the catalytic triad His-Tyr-Glu, (HYE). The histidine and tyrosine residues are responsible for the recognition and binding of NAD⁺,¹⁶ while the glutamate residue is necessary for PAR-polymerisation.¹⁷ Also present in the CD is a WGR motif, consisting of the residues (Trp, Gly, Arg), whose function is unclear.¹⁸

1.1.2 Plant PARP

Orthologues of mammalian PARP exist in plants. At least three types of PARP superfamily members are known in plants. The first evidence for poly(ADP-ribosylating) enzymes in plants was the finding of PARylated histones in *Nicotiana tabacum*¹⁹ and wheat nuclei.²⁰ Through genetic experiments²¹ and sequence similarities^{22,23}, plant PARP superfamily members were identified and data revealed that all land plants contain orthologues of *HsPARP1*. The best-investigated plant orthologue of *HsPARP1*, *Arabidopsis thaliana* PARP2 (*AtPARP2*, At2g31320), shares the same domain structure as *HsPARP1* and - as it is specific for all those members - shares the same catalytic triad histidine-tyrosine-glutamate (HYE).

Based on sequence similarity within the catalytic domains, some PARP have been identified as more closely related to *HsPARP3*.³ *HsPARP3* domain-related plant proteins are split into two groups. *AtPARP1*/APP (At4g02390) belongs to the first of those groups. It is also the first plant PARP that was cloned.²³ Members of this subgroup share a plant-specific domain structure which contains two N-terminal SAF-A/B, Acinus, and PIAS (SAP) domains that are involved in binding of nucleic acids²⁴ and protein localisation to the kinetichore during mitosis.²⁵ The PARP signature of this subgroup contains the conserved HYE motif. For *Zea mays* PARP1 (*ZmPARP1*)²⁶ and *AtPARP1*²² PARylation activity was demonstrated.

The second subgroup, to which *AtPARP3* (At5g22470) belongs, is more closely related to *HsPARP2*. In contrast to the first subgroup, the SAP domains are missing and the catalytic triad is disrupted. The histidine is replaced by a cysteine. And, while in seedless plants the

motif consists of CYE, in all angiosperms, this tyrosine residue is exchanged into CVE³ which indicates that NAD⁺ binding and consequently its enzymatic activity are unlikely. *At*PARP3 might have a function in the developmental stage of life cycle as *At*PARP3 is mainly expressed in developing seeds.²⁷

Proteins that are orthologous to the *Hs*PARP8 clade (*Hs*PARP6, 8 and 16) have been found in some green algae, moss and many fungi.³ Neither in humans nor in plants, have members of this clade been functionally characterised.

The catalytic domain of PARP has also been found in six further *Arabidopsis* genes. These genes encode proteins that are named Radical-induced Cell Death 1 (*At*RCD1, At1g32230) and the proteins Similar to RCD One 1-5 (*At*SRO1-5). Despite lacking poly(ADP-ribosylation activity in RCD1 and all SRO²⁸, it is speculated that these proteins have mono(ADP-ribosyl) transferase activities.¹⁰ Evidence is supporting the hypothesis that members of the SRO family are involved in the gene regulation at transcriptional or chromatin level. RCD1 and SRO1 have been shown to bind to transcription factors in yeast two-hybrid assays.^{29,30} These observations suggest similar roles of plant PARP family members to that known from human PARP.

Due to sequence analysis and the comparison of the domain composition between human and *Arabidopsis* PARP, the *Arabidopsis* PARP1 and PARP2 nomenclature has changed. The *Arabidopsis* PARP protein that is most similar to *Hs*PARP1 in terms of sequence similarity and sequence length was described as *At*PARP1 (At2g31320). With respect to domain structure conservation in comparison to *Hs*PARP1, the former *At*PARP2 (At4g02390) is most similar to *Hs*PARP1 and is therefore described as *At*PARP1. For the same reason, the former *At*PARP1 is now described as *At*PARP2. This will be the nomenclature used in this work.

1.1.3 The catalytic reaction of PARP

The catalytic reaction of PARP is exemplified on *Hs*PARP1 in Figure 1.1. The active site of PARP can be divided into a donor and an acceptor site. Positioned in the donor site, the substrate NAD⁺ donates an ADP-ribose unit to a nascent ADP-ribose chain, the acceptor molecule. Hereby, the pyridinium acts as leaving group, generating an electrophilic C₁ at the donor ribose. The first step of the polymer elongation reaction involves the concurrent binding of a molecule NAD⁺ in the donor site and the repositioning of an existing ADP-ribose chain in the acceptor site (Figure 1.1, A). The catalytic glutamate (Glu₉₈₈, *Hs*PARP1

numbering) plays a crucial role in the reaction. Firstly, one of the Glu₉₈₈ carboxyl oxygen atoms forms a hydrogen bond to the 2'-OH of the acceptor ribose. This polarises the acceptor oxygen and increases its nucleophilicity. Secondly, the nicotinamide ribose of NAD⁺ is bound in 3'-*endo* conformation in PARP, a conformation that is already close to an expected oxocarbenium transition state geometry. During the reaction, Glu₉₈₈ forms another hydrogen bond to the 2'-hydroxyl of the donor ADP-ribose which leads to a stabilisation of the oxocarbenium. The ADP-ribosyl transfer reaction takes place by a nucleophilic attack of the acceptor ribose 2'-OH on the C1'_N carbon of the donor ribose in which an $\alpha(1 \rightarrow 2)$ -glycosidic bond is formed, and nicotinamide is released.

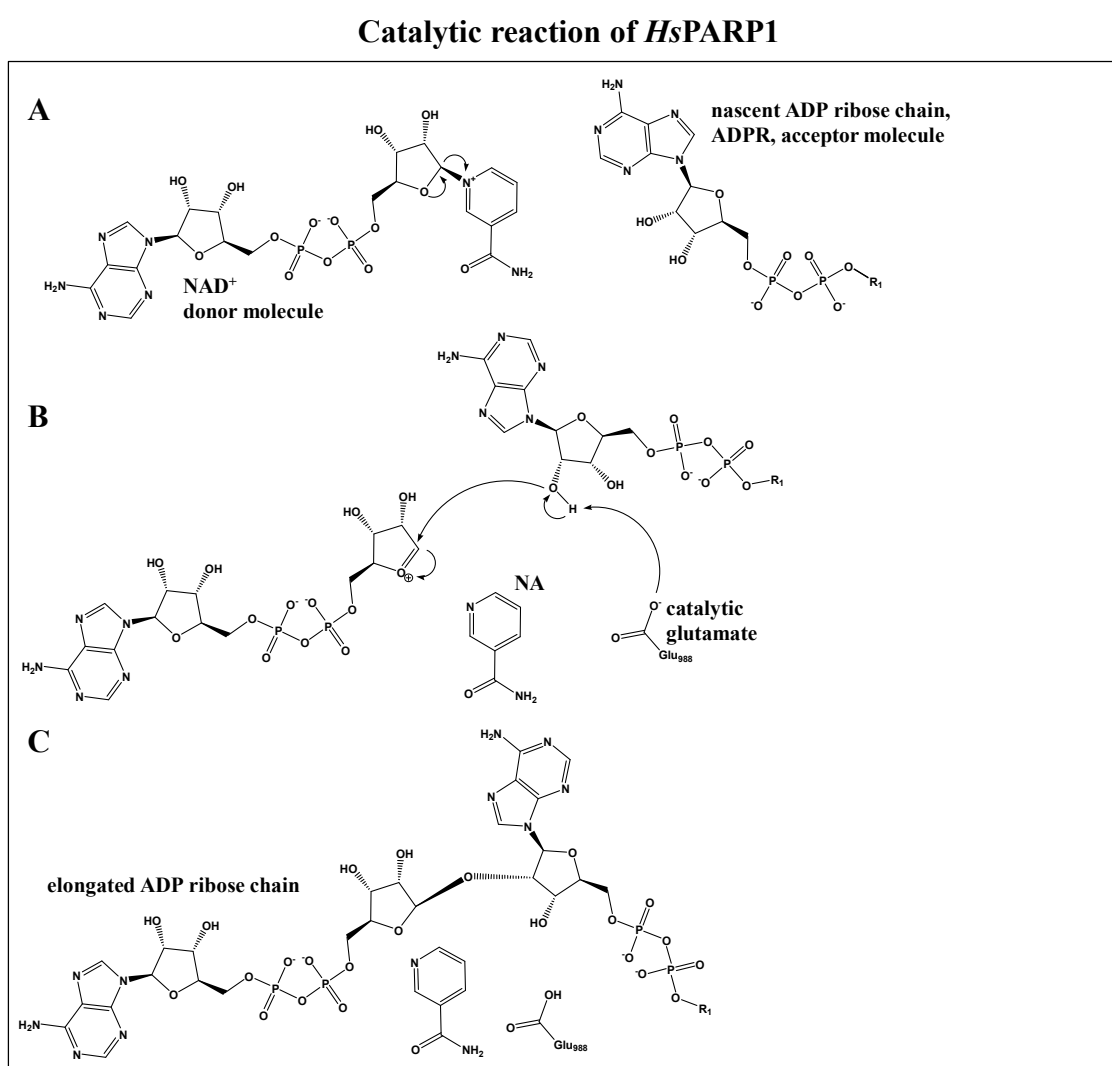


Figure 1.1: Catalytic reaction of PARP

A: ADPR chain approaches a bound molecule NAD⁺ B: nucleophilic attack of the acceptor ribose on the donor ribose, mediated through the catalytic glutamate 988, C: $\alpha(1 \rightarrow 2)$ glycosidic bond formation

The final process of the reaction is not clearly resolved. According to Ruf and co-workers, the catalytic reaction follows an S_N2 mechanism³¹ while an S_N1 mechanism was proposed by

Scheuring and Schramm based on an observed change in the hybridisation of the anomeric carbon.³² The hydrolysis of NAD^+ , resulting in the generation of the first ADP-ribose molecule, was investigated on a theoretical level using combined quantum mechanical/molecular mechanical (QM/MM) methods.³³ Based on their results, the authors concluded that the catalytic reaction is a concerted $\text{S}_{\text{N}}2$ reaction. Independent of the different conclusions concerning the character of the S_{N} -reaction, most studies agree that its transition state has an oxacarbenium character from which one can conclude that PARP's nucleophilic substitution reaction proceeds on the borderline of $\text{S}_{\text{N}}1$ and $\text{S}_{\text{N}}2$ mechanisms.

The reaction mechanism applies for the synthesis of a branched polymer, too. Here, the orientation of the acceptor molecule is reversed by a 180° rotation. Due to the internal symmetry of an ADP-ribose unit, the phosphate moiety positions in the same way as in the elongation reaction. In contrast to the elongation reaction, the glycosidic linkage is formed between the 2'-OH of the nicotinamide ribose and the anomeric $\text{C1}'_{\text{N}}$ of the donor ribose. The normal ratio of branching to elongation is 1:50. Evidence by Rolli and colleagues suggest that the asymmetry of PARP's acceptor site determines this ratio. The mutation Y986H in *Hs*PARP1 rendered the protein's acceptor site more symmetric which leads to an increased branching:elongation ratio towards 1:1.¹⁶ A schematic representation of the branching and elongation reaction, as proposed by Ruf and colleagues³¹, is shown in Figure 1.2.

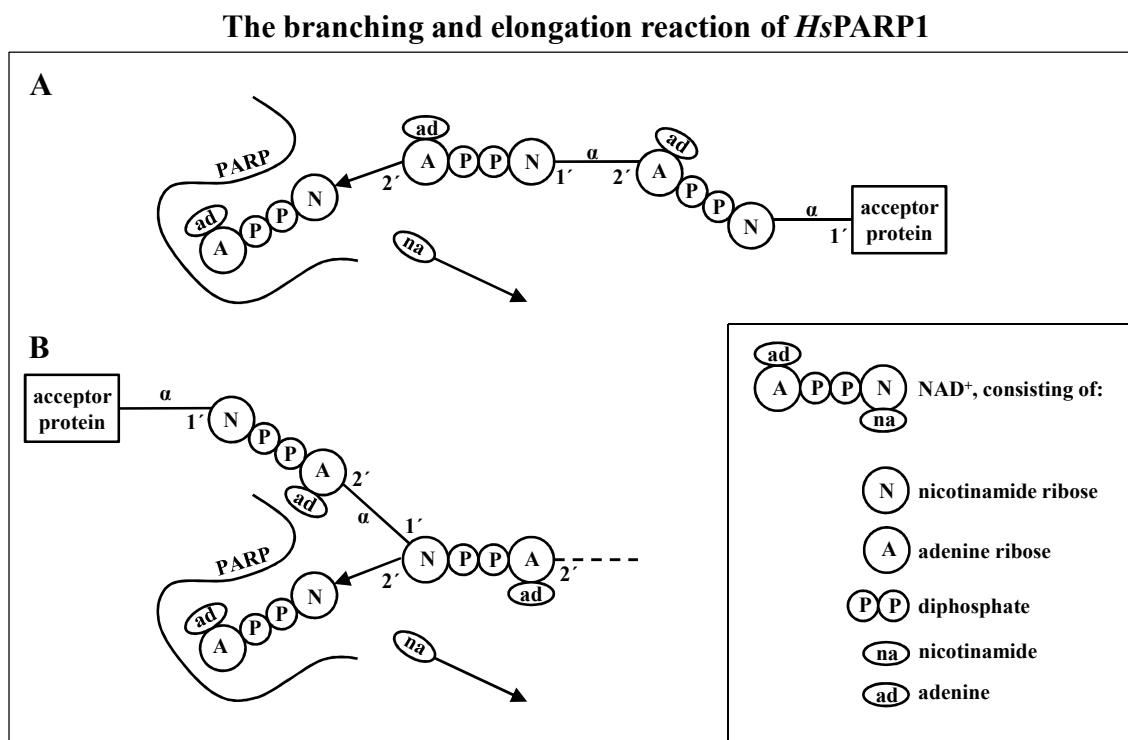


Figure 1.2: The proposed mechanism of the branching and elongation reaction of PARP Mechanisms as proposed by Ruf et al. (modified)³¹ A: The elongation reaction B: The branching reaction

1.2 *Hs*PARP1 as a therapeutic target

Poly(ADP-ribose) metabolism is stimulated by DNA damage and *Hs*PARP1 is involved in a DNA damage signalling network and DNA repair. *Hs*PARP1 and its counterpart Poly(ADP-ribose) glycohydrolase (*Hs*PARG) are the enzymes that contribute to the majority of poly(ADP-ribose) metabolism in human. PARP contributes to genomic integrity³⁴ since it is involved in different DNA repair mechanisms, as well as in telomer protection and DNA damage signalling that can lead to cell cycle survival, cell cycle arrest, cell transformation or cell death. *Hs*PARP1 modulates chromatin structure where it interacts with histones H1-H4³⁵, guides chromatin decondensation and transcriptional activation through poly(ADP-ribosylation).³⁶ PARP has several interaction partners that are involved in DNA repair. Among those are DNA-Ligase III³⁷, DNA-Polymerase β ³⁸, X-ray repair cross-complementing 1 (XRCC1)³⁹ and PARP2.⁴⁰ It also interacts with transcription factors among whose are nuclear factor kappa-light-chain-enhancer of activated B cells (NF- κ B)⁴¹ and p53.⁴² PARP participates in replication via interactions with DNA-Ligase I and DNA-Polymerase α .⁴³

1.2.1 *Hs*PARP1 and DNA repair

Damages on the DNA arise from endogeneous and exogeneous factors as reactive oxygen species (ROS), alkylating and cross-linking agents, non-enzymatic hydrolysis of the phosphodiester backbone of nucleic acids and electromagnetic radiation. Resulting DNA damage can be divided into three groups that, depending on the severity of DNA damage, are repaired by different DNA repair mechanisms.⁴⁴ Minor damage like oxidation or methylation of DNA bases or DNA single strand breaks are removed by the Base Excision Repair (BER) or Single Strand Break Repair (SSBR) systems.^{45,46} Moderate DNA damage like dimerised pyrimidins is eliminated by the Nucleotid Excision Repair (NER) system.⁴⁷ Major damage like DNA double-strand breaks (DSB) are corrected by Nonhomologous End Joining (NHEJ) or Homologous Recombination (HR) systems^{44,48} while DNA replication errors are adjusted by the system of Mismatch Repair (MMR)⁴⁹

Several studies showed an involvement of PARP in the SSBR and BER systems. Once PARP detects a single-strand break, it binds on the location of the damage and autoPARylates itself. The PARylation induces the recruitment of XRCC1. The single-strand break is subsequently repaired and ligated by the proteins polynucleotide kinase/phosphate, DNA-polymerase β and DNA-ligase III.^{45,50}

In the BER system, specific DNA glycolases identify and cleave the modified base resulting in apurinic or apyriminic sites (AP sites) in the DNA. The site is subsequently cleaved by AP-endonuclease 1 preparing the site to be repaired by DNA-polymerase β and ligated DNA-ligase III. The presence of PARP was proposed to not being essential since PARP is not directly involved.⁵¹ But findings with 3-aminobenzamide (3AB) showed improved efficiency in BER when PARylation was activated. An involvement of PARP in BER was also proposed through experiments in several mouse-models.⁵² The theory of PARPs's role in BER was further supported by interaction studies in which the interaction of PARP1 with XRCC1⁵³, DNA-polymerase β ⁵⁴ and DNA-ligase III³⁷ was shown. This indicates an indirect participation of PARP1 in the BER. In 2004, a different model has been developed that discusses at which stage PARP participates in BER and establishes protein-protein interactions.⁵⁵ The last step of both the BER and SSBR are equivalent since once the damaged site of DNA is repaired, PAR polymers become degraded by PARG und PAR-bound proteins disengage. After the automodification status of PARP1 and PARP2 is reversed, the proteins are enabled for being involved in another cycle of DNA damage repair.

1.2.2 HsPARP1 and synthetic lethality

Two genes X and Y are synthetically lethal if mutations in one of the genes alone are viable but mutations in both genes occurring simultaneously are lethal. The concept of synthetic lethality was proposed in the 1990s as an alternative to select new anticancer drug targets.⁵⁶ If X and Y are synthetic lethal, than inhibitors of Y should selectively inhibit or kill cancer cells having mutant X. An extraordinary advantage of this concept is that even a complete inhibition of Y would have no effect on normal cells and even partial inhibition of Y would kill cancer cells having mutations in X. Human PARP1 was linked to synthetic lethality in 2005, when two independent groups showed that breast cancer associated genes 1 and 2 (*BRCA1* and *BRCA2*) -deficient cell lines are sensitive toward *HsPARP1* inhibitors.^{57,58} *BRCA1* and *BRCA2* have been characterized as tumour suppressor genes.^{59,60} They are involved in HR, a process involved in the repair of DNA double strand breaks.⁶¹ The prevalence of DNA single strand breaks caused by *HsPARP1* inhibitors will lead to DSB via replication fork collapse.⁶²

Chromosomal aberrations and genome instability are consequences of increased DSBs in HR-deficient cell types that eventually lead to cell death. The concept of synthetic lethality in connection with PARP inhibitors would therefore effectively kill tumor cells that have

deficiencies in BRCA1/2 while not affecting normal cells. This is of interest because carriers of heterozygous BRCA1 or BRCA2 mutations are prone to develop breast cancer and are also predisposed to ovarian, prostate and pancreatic cancer. In principle, any cell line that lacks the ability for HR could be tested for PARP inhibitors sensitivity. Tumour types with inactivated HR pathways are described as “BRCAness”.

The concept of synthetic lethality and its relation to PARP inhibition

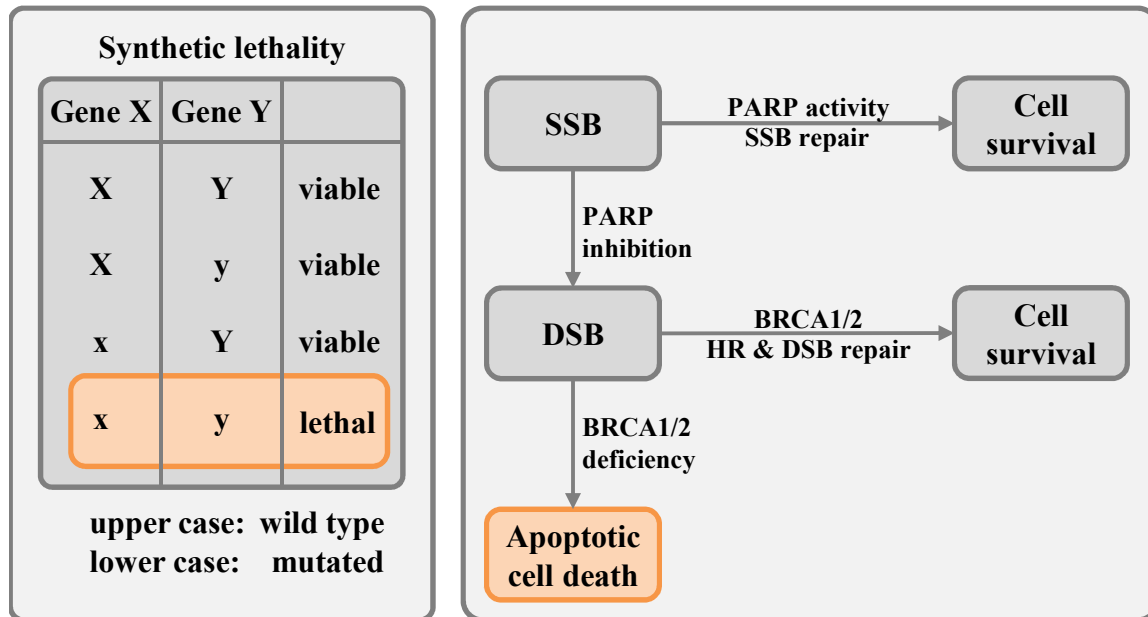


Figure 1.3: Synthetic lethality and PARP inhibition

Abbreviations: SSB: single strand break, DSB: double strand break, BRCA: breast cancer associated gene, HR: homologous recombination

1.2.3 HsPARP1 and ischemia

The state in which a tissue suffers from restricted blood supply is known as ischemia. As a consequence, there is a deficit in molecular oxygen supply (hypoxia) in the damaged tissue that can lead to impaired cellular functions and ultimately to cell death. Although the reperfusion of ischemic tissue with oxygenated blood should reinstate normal physiological functions, the reperfusion also contributes to the overall injury that is caused by Ischemia-reperfusion (IR). This phenomenon is called “reperfusion injury” (RI).

Excessive activation of PARP can lead to a rapid consumption of cellular NAD^+ pools. NAD^+ depletion leads to a decrease in ATP pools as well, as NAD^+ acts as an electron carrier in mitochondrial respiratory chain. Nicotinamide, the released by-product during NAD^+ -cleavage by PARP, can be recycled back to NAD^+ . This process again requires ATP. The

rapid fall of ATP pools upon continuous PARP activation via two different mechanisms can finally lead to cell death.⁶³

During ischemia-reperfusion, oxygen-derived radicals like superoxide anions ($O_2^{\bullet-}$) and hydroxyl radicals (OH^{\bullet}) can cause DNA strand breaks. Also, the nitrogen-derived radical nitric oxide (NO^{\bullet}) reacts with superoxide anions and produces peroxynitrite ($ONOO^-$) during IR. Peroxynitrite itself causes DNA strand breaks, too, that lead to PARP activation and cell death. In 1997 it could be shown that PARP1 knock-out mice displayed more than 60% reduction of damaged tissue in an animal model of stroke. This gave evidence that PARP inhibitors could reduce the amount of damaged brain tissue in stroke patients and therefore displaying therapeutic benefits.

1.3 PARP inhibitors

1.3.1 Development of *Hs*PARP inhibitors

Since PARP is involved in DNA repair, it has been seen that inhibition of DNA repair via PARP inhibition leads to sensitization of tumor cells when used in combination with chemo- and radiotherapy or in specific genetic backgrounds. Alkylating agents like temozolomide (TMZ), camptothecins and radiation are widely used in therapies and produce SSB which cannot efficiently be repaired with inhibited or disrupted PARP. The first enzyme-selective PARP1 inhibitor was 3AB⁶⁴ which in the same year was shown to enhance cytotoxicity caused by preventing the rejoining of DNA strand breaks by the alkylating agent dimethyl sulfate and increased its toxicity in L1210 mouse leukemia lymphoblast cells.⁶⁵ Despite 3AB being a simple analogue of NA (1, Figure 1.4) and a weak and unselective PARP inhibitor (IC_{50} of $30\mu M$ ⁶⁶), the results of that study led to the development of more potent inhibitors having isoquinolinone⁶⁷, quinazolinone or phenantridinone core structures. Those core structures were used as lead compounds with potencies that were sufficient to use them in pre-clinical trials.⁶⁸

Rational drug design was further supported by crystallographic studies of the catalytic domains of PARP that were deposited in the Protein Data Bank (PDB).⁶⁹ Those PARP-domain structures were derived from *Hs*PARP1 (e.g. PDB entry 1UK0⁷⁰), *Hs*PARP2 (e.g. PDB entry 3KCZ⁷¹) and *Gallus gallus* PARP1 (*Gg*PARP1, e.g. PDB entry 2PAX⁷²), confirming residues responsible for inhibitor binding and suggesting a common binding mode

of PARP's substrate NAD^+ . This led to the development of more potent compounds having low toxicity and that were active in combination studies with anticancer chemotherapies in xenograft models. For example, antitumor activity of TMZ, irinotecan and cisplatin against tumour xenograft in mouse was increased by CEP-6800⁷³ (2, Figure 1.4), antitumor efficacy of TMZ against melanoma, glioblastoma multiforme, and lymphoma growing in the mouse brain was enhanced by GPI 15427⁷⁴ (3, Figure 1.4) and an improved therapeutic index was found with AG14361 (4, Figure 1.4) in combination with TMZ, irinocetan and radiation in a human colon tumor xenograft model.⁷⁵ The increased antitumor activity of TMZ in combination with a PARP inhibitor is caused by hindering the BER which removes methylpurine species that are generated by TMZ.⁷⁶

Examples of developed *Hs*PARP1 inhibitors

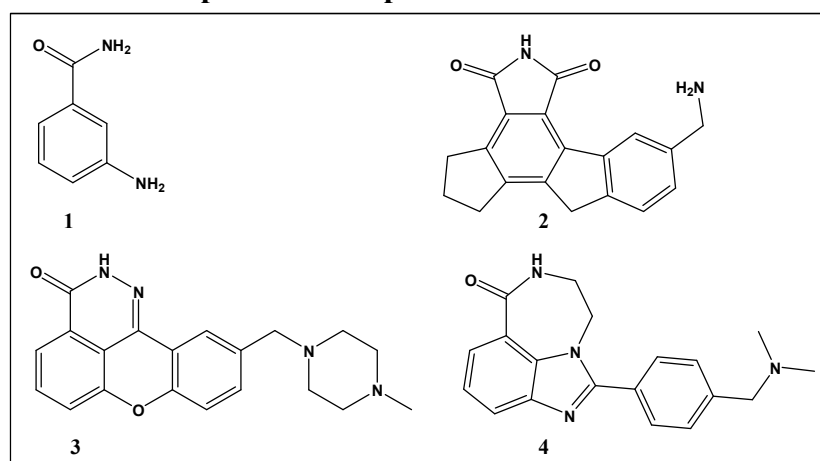


Figure 1.4: Examples of developed PARP inhibitors
Structures: 1 3AB, 2: CEP-6800, 3: GPI 15427, 4: AG14361

1.3.2 *Hs*PARP inhibitors in clinical trials

Due to PARP's roles in DNA repair, in pathological conditions that involve restricted blood flow and the findings of connections between PARP inhibitors and *BRCA1/2*-deficient cell lines, 129 PARP-associated clinical trials have been enrolled, are in progress or have already been finished with published results (www.clinicaltrials.gov, accessed on 29.12.2014)⁷⁷. In the majority of these clinical trials, PARP inhibitors are used in a cancer setting that either combine the PARP inhibitor with standard chemotherapeutic protocols or test a PARP inhibitor as monotherapy to treat tumours that are defective in their DNA repair machinery. Since 2003, 11 different compounds underwent clinical trials (Table 1.1). While aspects like metabolic stability or bioavailability are limitations for inhibitors to enter the market, further

challenges such as resistance to PARP inhibitors and polypharmacology of PARP inhibitors have recently been identified in drug development.

The first clinical trial of a PARP inhibitor was carried out in 2003 with the tricyclic indole inhibitor Rucaparib (AG-014699, PF-01367338) in combination with TMZ in patients with advanced solid tumors (5, Figure 1.5).⁷⁸ Rucaparib was selected from different series of benzimidazole carboxamides as a candidate having promising inhibitory effects (K_i 1.4 nM) and improved solubility.⁷⁹ It was successfully used in clinical trials phase II to treat patients with advanced metastatic melanoma⁸⁰ and is now being used in phase II as a stand-alone therapy for advanced breast or ovarian cancer in patients having *BRCA1/2* deficiencies.

Examples of *Hs*PARP1 inhibitors having entered clinical trials

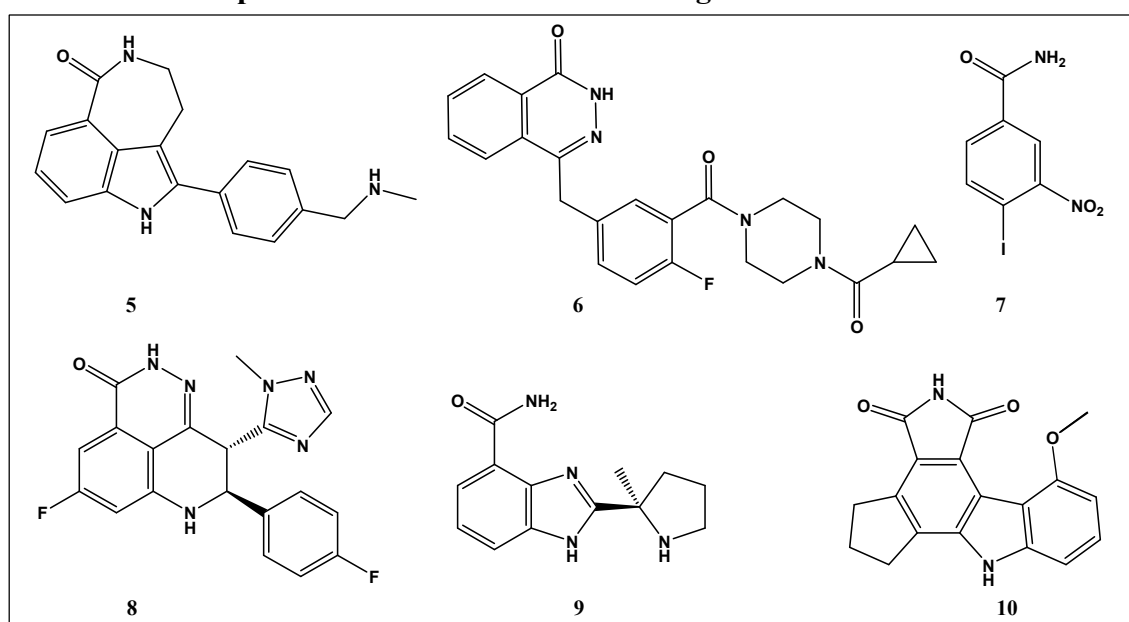


Figure 1.5: Examples of human PARP1 inhibitors having entered clinical trials

Structures 5: Rucaparib (AG-014699), 6: Olaparib (AZD-2281), 7: Iniparib (BSI-201), 8: Talazoparib (BMN-673), 9: Veliparib (ABT-888), 10: CEP-8983

The PARP inhibitor CEP-8983 (10, Figure 1.5) showed high potency (K_i 20 nM), but was of limited solubility.⁸¹ The problem was solved by developing CEP-9722 (structure not shown) which has improved solubility and acts as a pro-drug of CEP-8983. Promising results in pre-clinical trials indicated CEP-9722 as a chemosensitising agent.⁸¹ CEP-9722 is now used in three phase I clinical trials, either used as single-agent therapy used in patients having advanced solid cancer or as a combination therapy together with TMZ or gemcitabine or cisplatin in patients with metastatic solid tumours or mantle cell lymphoma.

Olaparib (AZD-2281) belongs to the PARP inhibitor class of phthalazinones.⁸²⁻⁸⁴ Structural improvements led to optimized inhibition potency, metabolic stability, increased solubility and oral bioavailability. Olaparib (6, Figure 1.5) has entered clinical trials I, II and III, in which it is used as a single-agent or in combination with chemotherapeutic drugs after its potency was shown in pre-clinical trials.⁵⁸ But concerns arised as restorations of BRCA-functions by secondary mutations as well as induction of P-glycoprotein transporters led to chemoresistance.⁸⁵ Patients showing resistance to Olaparib also showed secondary *BRCA*-mutations. Those mutations restored DNA repair in tumour cells.⁸⁶ The Olaparib-related compound AZD-2461) showed growth inhibition of drug-resistant clones in long-term application. It has now entered clinical trial phase I in which its safety in patients with refractory solid tumours is assessed.

Table 1.1: HsPARP1 inhibitors in clinical trials

Drug		Clinical trial phases					
Drug name	synonym	Ph I	Ph I/II	Ph II	Ph III	unk.	Σ
ABT-888	Veliparib	23	4	10	2	0	39
AZD-2281	Olaparib	19	2	13	2	1	37
BSI-201	Iniparib	5	1	10	2	1	19
BMN-673	Talazoparib	6 *	2	3	1	0	12
AG-014699	Rucaparib	1	1	4	1	0	7
MK-4827	Niraparib	3	0	1	2	0	6
CEP-9722		2	1	0	0	0	3
INO-1001		1	0	2	0	0	3
AZD-2461		1	0	0	0	0	1
E7016		1	0	0	0	0	1
E7449		0	1	0	0	0	1
	Σ	62	12	43	10	2	129

Ph = Phase, unk.=unknown; * one existing trial in Phase 0 included
data from <http://www.clinicaltrials.gov>, accessed: 29.12.2014

A promising PARP inhibitor that was dismissed later on is Iniparib (BSI-201, 7 in Figure 1.5). It was claimed to non-competitively inhibit PARP, but was later shown to modify a broad range of cysteine-containing proteins.^{87,88} It was the first PARP inhibitor entering clinical phase III to be tested in patients having breast cancer and squamous lung cancer. Due to discouraging results, Sanofi announced to end the research in early 2013.⁸⁹

The PARP inhibitor being used in clinical trials so far is Veliparib, a benzimidazole-carboxamide derivative. First results in preclinical tumor models allowed for testing in clinical trials phase I⁹⁰ and later on in phases II.⁹¹ Currently, there are two phase III trials recruiting in which the effects of Veliparib are examined together with Paclitaxel and Carboplatin treatment in more than 1000 patients (NCT02163694 and NCT02106546).

1.4 The role of poly (ADP-ribosylation) in plants

1.4.1 Plant PARP

The functions of PARP in human are investigated since the 1960s.² It is now well established that human PARP are involved in DNA repair, replication and transcription (1.1.1). The functions of PARP may be inferred to be conserved between human and plant PARP due to the high degree of conservation of domain structures and at amino acid level. PARP are DNA break sensors and DNA repair signalling molecules. They are first responders to sites of DNA breaks. Findings for *Arabidopsis thaliana*, where *AtPARP1* and *AtPARP2* mRNAs accumulate quickly both after γ -radiation and the accumulation of reactive oxygen species (ROS), support evidence of similar functions of plant PARP.⁹² Overexpression of *AtPARP2* led to decreased levels of ROS-induced DNA nicks.⁹³ *AtPARP1* and *AtPARP2* expressions rise in genetic backgrounds that are characterised by increased DNA damage or replication stress.^{22,94,95} In stem cells and rapidly dividing tissues, *AtPARP1* and *AtPARP2* are escalated as well which indicates PARP's involvement in genome integrity. The expression is induced by radiation^{92,96} or genotoxic stress.⁹⁷ *AtPARP2* expression is also increased by oxidative stress and salinity^{92,97,98}, similar changes in *AtPARP3* expression were observed upon treatment with *N,N'*-dimethyl-4,4'-bipyridinium dichloride (Paraquat), induced salinity, high light or drought stress.⁹⁸ Abiotic stresses that lead to oxidative stress, induce PARP activity that is responsible for diminishing of cellular NAD^+ and consequently ATP pools. Inhibition of PARP may minimise the depletion of NAD^+ and ATP pools, resulting in enhanced tolerance against these stresses. Similar effects were observed by downregulation of *AtPARP1* and *AtPARP2* by RNAi, where NAD^+ consumption and stress-induced PARylation were reduced and ATP pools retain at higher levels. This led to decelerated ROS accumulation and increased stress tolerance.⁹⁹

Like mammalian PARP, plant PARP are implicated in programmed cell death (PCD). In soybean cells, PARP are activated and cellular levels of NAD^+ decline upon induction of oxidative stress. PARP inhibition or down regulation might delay PCD.⁹³ Further evidence of the connection between PARP and PCD is an improved resistance of soybean cells to mild oxidative stress after *AtPARP1* overexpression in soybean.⁹³

Furthermore, there are studies that link PARP activity with the plant hormone abscisic acid (ABA). Increased levels of cyclic ADP-ribose (cADPR), which is synthesised from NAD^+ , are observed in *PARP*-deficient plants.¹⁰⁰ Together with ABA and Ca^{2+} , cADPR acts as a

second messenger. Changes in ABA levels due to abiotic stress are observed before changes in gene expression.¹⁰¹ In *Arabidopsis*, more than 100 ABA-responsive genes can be induced by increased levels of cADPR.¹⁰² *AtPARP*-deficient plants, since being unable to cleave NAD^+ , could provide more NAD^+ for cADPR production that finally leads to enhanced stress tolerance by improved production of ABA-regulated stress response genes.

1.4.2 Plant PARG

The transfer of ADP-ribose moieties from NAD^+ to target proteins is reversible. Proteins that hydrolyse PAR polymers and generate free ADP-ribose are called poly(ADP-ribose) glycohydrolases (PARG). By catalysing this reaction, cellular pools of unbound ADP-ribose are increased. Unbound ADP-ribose is a known cell death signal.¹⁰³ Since ADP-ribose can also be cleaved from target proteins, it enables them for further ADP-ribosylation. The function of PARG in counteracting or contributing to the impacts of PARP activity is context-dependent. In animal systems, PARG plays a crucial role in cell death embryonic development¹⁰⁴ and DNA repair.¹⁰⁵ Most animal genomes contain one single *PARG* gene which, when knocked out in mice¹⁰⁶ or *Drosophila*¹⁰⁴, results in lethality due to the accumulation of toxic PAR polymers.

Arabidopsis encodes two adjacent *PARG* genes (At2g31865 and At2g31870) and one pseudo gene (At2g31860). Some plants species (among which are *Oryza sativa*, poplar and *Zea mays*) are predicted to encode for two or more *PARG* genes, whereas other plant species (like *Ricinus communis* and *Sorghum bicolor*) are predicted to encode for one gene. PARG1 enzymatic activity was shown in *Arabidopsis*. Higher concentrations of ADP-ribose polymers in PARG1-deficient plants compared to wild-type plants have been observed in *Arabidopsis thaliana*.¹⁰⁷ Although plant PARG is not as well investigated as PARP, evidence suggests that PARG are involved in regulation of circadian clock in *Arabidopsis*. PARG1-mutated plants show an increased leaf movement and cause early flowering under short and long days. It has also been seen that PARG1-mutated plants lengthen the period of all known circadian clock-controlled genes.⁸⁸

1.4.3 Plant NUDX

ADP-ribose-specific Nucleoside Diphosphate linked to X hydrolases (NUDX) are proteins that degrade free ADP-ribose into adenosine monophosphate (AMP) and ribose-5-phosphate (R5P). Free ADP-ribose can non-enzymatically mono(ADP-ribosyl)ate proteins and is highly

reactive. High levels of free ADP-ribose are toxic. NUDX activities contribute to NAD^+ maintenance by supplying a source for ATP during the cleavage of ADP-ribose. NUDX are key proteins in re-establishing the cells energy levels.^{98,108,109}

Roles of PARP, PARG and NUDX proteins in abiotic stress response

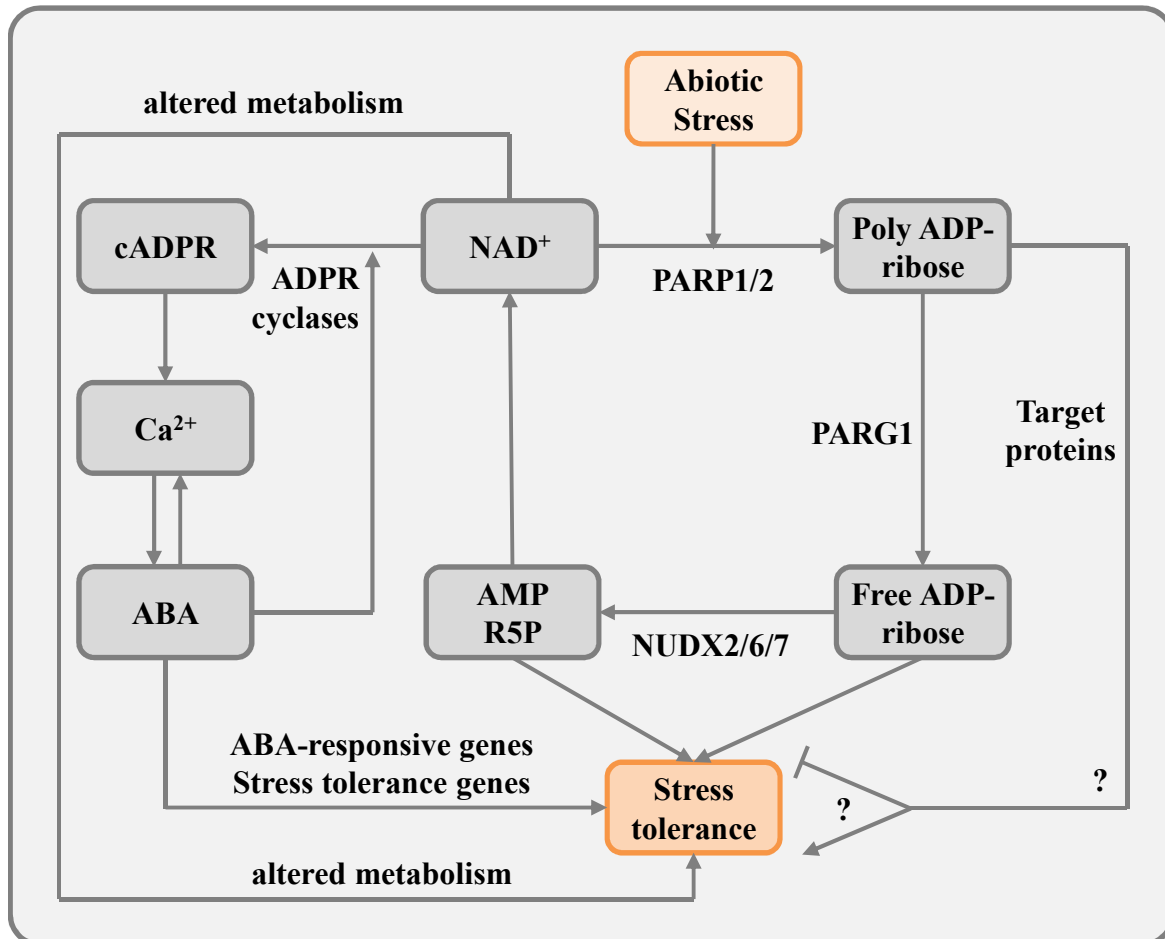


Figure 1.6: Interplay between PARP, PARG and NUDX proteins in abiotic stress
Scheme modified from Briggs¹¹⁰

In *Arabidopsis*, there are 27 genes that encode proteins (*AtNUDX1-AtNUDX27*) having a NUDX box domain which is identified by the motif $\text{GX}_3\text{EX}_7\text{REUXEEXGU}$.^{98,109} *AtNUDX1-AtNUDX11* target the cytosol, and among those, *AtNUDX7* seems the most prevalent NADH and ADP-ribose pyrophosphatase in *Arabidopsis* cells.^{111,112} *AtNUDX12-AtNUDX18* target mitochondria and *AtNUDX19-AtNUDX24* target chloroplasts. It has been shown that *AtNUDX2*, *AtNUDX6*, *AtNUDX7* and *AtNUDX10* hydrolyse ADP-ribose and NADH to AMP *in vitro*, while substrates such as 8-oxo-dGTP, dNTPs, NADH, CoA and FAD can be hydrolysed, too.

AtNUDX7 gene expression is upregulated by virulent and avirulent pathogens while a reduction of hypersensitive response to an avirulent pathogen was observed in *Arabidopsis* knock-out *nudx7*.^{113,114} *AtNUDX7* has also been linked to abiotic stress because environmental stresses cause microscopic necrotic lesions, ROS are accumulated and *Atnudx7* mutants are stunted.^{112,115}

The interplay between PARP, PARG and NUDX in *Arabidopsis thaliana* is schematically displayed in Figure 1.6.

1.5 Virtual screening in Lead Discovery

Lead discovery and testing of lead compounds in the pharmaceutical and agrochemical industry are different in both industries. A question in lead discovery lies in identifying a target on which a lead acts. Furthermore, the mode of action of the lead needs to be clarified before a lead or drug can enter the market. A pharmaceutical lead is usually defined as a compound having a modulating (e.g. inhibitory) activity against an enzyme or receptor *in vitro*. Having the target isolated, purified (or even crystallized) the activity of a lead in a first step is measured *in vitro*. After optimising structural properties of the lead and predicting potential side effects, as well as pharmacokinetic and pharmacodynamic properties, it will have to be successfully tested in animal models before it can enter clinical trials.

In contrast to that, an agrochemical lead is commonly defined as a compound having desired activity *in vivo*. Testing agrochemicals *in vivo* early in the discovery process (e.g. under glass house conditions or specific high-throughput-screening (HTS) set-ups) has potential benefits. Desired but also adverse effects of the drug become obvious very early. But as the lead (in general) is intended to act on a specific target, candidates that exert *in vivo* activity could potentially function on other targets, too. The use of HTS techniques for *in vivo* experiments can be used as a first filter in the discovery process, before the actual mode of action of leads is further investigated. Transferring glass house conditions to field conditions and characterisation of biochemical mode of action is a further obstacle that narrows the number of compounds.

Mathematical modelling and *in silico* screening techniques can help guiding the lead discovery. During the last decades, techniques that predict *in vitro* activities of lead candidates on the target structure or modelling tools to improve lead properties or help in

understanding the mode of action of a drug *in silico* have been developed and successfully applied. Some of them have been shown to link pharmaceutical and agrochemical industry. A famous example is Nitisinon which inhibits an enzyme involved in tyrosine catabolism. It is now used as a drug to treat the symptoms of the rare diseases hereditary tyrosinemia type 1¹¹⁶ and alkaptonuria.¹¹⁷

1.5.1 Virtual screening for human PARP inhibitors

Most of the leads as a starting point for the development of potent PARP inhibitors were identified by HTS or rational drug design. Companies invest heavily in the development of PARP inhibition assays and HTS systems. So have precursors of Olaparib at KuDOS Pharmaceuticals been identified and measured with a flash plate assay system which was developed by this company for this very purpose.¹¹⁸ While HTS for PARP inhibitors account for the majority of currently developed PARP inhibitors, molecular modelling and virtual screening (VS) techniques became relevant since the 1990s. In 1991, the NAD⁺ binding sites of ADP-ribosylating toxins, including *Pseudomonas aeruginosa* exotoxin A (ETA) and Diphtheria Toxin (DT), were computationally modelled and compared.¹¹⁹ Structural similarities were further investigated in 1994 and the role of a conserved glutamate was already discussed, two years before the crystal structures of DT or GgPARP1 were published and released.¹²⁰ Both the crystal structures of DT¹²¹ and GgPARP1¹²² were solved two years later and the proposed importance of a catalytic glutamate and also histidine residues could be verified by structure determination. The binding mode of NAD⁺ in GgPARP1 was modelled in 1998 and mechanisms for the branched and elongation reaction of PARylation were proposed.^{31,72} Although for ADP-ribosylating toxins, there are now crystal structures available for ETA (PDB entry 3B8H)¹²³, DT (PDB entry 1TOX)¹²¹ and Cholera Toxin (PDB entry 2A5F)¹²⁴, the binding mode of NAD⁺ in PARP could only be modelled so far. A common structural binding motif for NAD⁺ in poly(ADP-ribose) polymerases and ADP-ribosylating toxins was proposed by Lee in 2010. The authors suggested a “scorpion motif” which is determined by a conserved YX₁₀Y sequence. This motif comprises a conserved loop having only small C_α RMSD values upon superpositioning and that is responsible for recognising NAD⁺. The authors used this structure motif and docking by fitting and model NAD⁺ into PARP enzymes with more structural confidence.¹²⁵

In 1998, more crystal structures of PARP catalytic domains were solved with different classes of GgPARP1 inhibitors such as 4-amino-benzo[de]isoquinoline-1,3-dione (4AN in PDB entry

2PAX), 3-methoxybenzamide (3MB in PDB entry 3PAX), and 8-hydroxy-2-methyl-3-hydroquinazolin-4-one (NU1 in PDB entry 4PAX).⁷² Also, two crystal structures, PDB entries 1A26³¹ and 2PAW⁷², of the catalytic domain of *Gg*PARP1 were crystallised; both having no inhibitor bound in the donor site in which NAD^+ is cleaved into NA and ADP-ribose. In these two structures, water molecules are present in the donor site instead of NA-mimicking inhibitors, giving insights into conserved water molecules in PARP and their role in the catalytic reaction. Investigations of the role of conserved water molecules in the PARP active site were used to analyse the contribution of water molecules for protein-ligand interactions.¹²⁶

The first quantitative structure-activity relationship (QSAR) has been carried out in 2001 with sets of 46 known *Hs*PARP1 inhibitors.¹²⁷ Docking studies were conducted prior to generating surface maps of the active sites from which descriptors that are related to entropy and enthalpy contributions were derived. These descriptors were used to perform QSAR analysis based on multiple linear regression (MLR) that yielded remarkable results. Prediction with an external test set of four compounds revealed excellent results ($r^2 = 0.795$, $q^2 = 0.720$) that enabled combined docking and QSAR methodologies to be used for rational design for new PARP inhibitors. Further QSAR modelling was used to derive precise models of 2-(1-propylpiperidin-4-yl)-1*H*-benzimidazole-4-carboxamid activities.¹²⁸ Here, genetic algorithm-multiple linear regression (GA-MLR) was used to derive models with high predictive power ($R^2 = 0.935$, $Q^2_{LOO} = 0.894$) for 34 structures. Benzimidazole caboxamide derivatives were also investigated with combined docking, QSAR, CoMFA¹²⁹ and CoMSIA¹³⁰ studies, in which highly predictive models (for CoMFA $r^2 = 0.899$, $q^2 = 0.712$ and for CoMSIA $r^2 = 0.889$, $q^2 = 0.744$) could be derived based on a much larger data set with 145 structures.¹³¹ Application of MLR and a feed forward neural network (FFNN) were used to study the behaviour of 30 phthalazinone derivatives. Using MLR ($r^2 = 0.766$, $r^2_{cv} = 0.694$) and FFNN ($RMS_{test} = 0.32$) with 14 descriptors allowed predictions of new phthalazinone analogues.¹³²

Molecular modelling was also used to support the design and synthesis of novel 4*H*-thieno[2,3-*c*] isoquinolin-5-one derivatives.¹³³ Structures of synthesised compounds were docked into the active site and frontier orbitals and electrostatic potentials were calculated for two structures whose activity could not be explained by visual inspection. First *in silico* investigations of selective PARP inhibitors were conducted by Ishida and colleagues in 2006.¹³⁴ To explain selectivity of synthesised inhibitors, they used the crystal structure of

*Hs*PARP1⁷⁰ and a homology model of *Hs*PARP2, based on the crystal structures of murine PARP2 (PDB entry 1GS0¹³⁵) for structural investigations. The selectivity of human PARP1/2 inhibitors was further investigated *in silico* by Novikov and co-workers in 2009.¹³⁶ They used *Hs*PARP1/2 as a case study to evaluate a developed docking program Lead Finder.¹³⁷ Lead Finder was used for virtual screening, binding energy calculations and for predicting selectivity between *Hs*PARP1 and *Hs*PARP2 inhibitors. Their test set consisted of 142 (selective) *Hs*PARP1/2 inhibitors and it could be shown that binding affinity could be predicted in an acceptable manner based on docking procedures. Another application in which PARP inhibitors were used in the context of virtual screening was published in 2010. There, in the group of Exner, a flexible ligand alignment technique for rapid superpositioning of ligands similar to pharmacophore searches was developed. Since PARP inhibitors share common chemical features as binding motif, test sets of known human PARP1 inhibitors were used to verify the potency of this approach.¹³⁸

1.5.2 Virtual screening in agrochemistry

One aim of agrochemistry is the development of new insecticides, fungicides and herbicides. The latter ones are compounds that affect the growing behaviour of weed by effectively and selectively inhibiting specific enzymes. Since potential plant PARP inhibitors will probably be applied as potential growth-affecting agrochemicals, in this section it will be focussed on the application of molecular modelling techniques on herbicides.

Around 300 compounds that act as herbicides have entered the market, all of them acting on less than 30 sites of action.¹³⁹ According to the Herbicide Resistance Action Committee (HRAC), compounds are classified into distinct sites of action and into chemical families within each site of action.¹⁴⁰ The classes mainly represent sites of photosynthesis and amino acid synthesis. Others act on disruption of cell division, seedling growth or synthetic auxins. A modified listing from HRAC representing the herbicide targets is shown in Table 1.2.

About half of all marketed herbicides act on three targets, namely acetolactate synthase (ALS), also known as acetohydroxy acid synthase (AHAS)^{141,142}, photosystem II (PSII), and protoporphyrinogen oxidase (PPO, EC 1.3.3.4). Also, only about 3% of all herbicides account for half of market share. The fact that the most prominent herbicides act on a few targets indicates that there is a need for identification of new targets. Out of the ~28.000 genes identified in *Arabidopsis*, about 20% of them are annotated as enzymes. Among the enzymes

that are identified as targets for herbicides, virtual screening techniques have been developed to screen for inhibitors or potentiate the activity of existing ones.

Table 1.2: Classification of Herbicides according to HRAC

class	Site of action	Chemical family (examples)
A	ACCase inhibition	cyclohexanedione, phenylpyrazoline
B	ALS or AHAS inhibition	triazolopyrimidine, imidazolinone
C1-C3	photosynthesis Inhibition at photosystem II	triazine, triazinone, urea, uracile, nitrile
D	Photosystem-I-electron diversion	bipyridilium
E	PPO inhibition	thia- and oxadiazole, triazolinone,
F1-F3	Bleaching: carotenoid biosynthesis at PDS, 4-HPPD and “unknown target” inhibition	pyridazinone, triketone, triazole, urea
G	EPSP synthase inhibition	glycine
H	glutamine synthetase inhibition	phosphinic acid
I	DHP synthase inhibition	carbamate
K1-K3	microtubule assembly or organisation, mitosis, VLCFAs (cell division) inhibition	dinitroaniline, benzamide, benzoic acid, carbamate, acetamide, tetrazolinone
L	cell wall (cellulose) synthesis inhibition	nitrile, benzamide, triazolocarboxamide
M	uncoupling (membrane disruption)	dinitrophenol
N	lipid synthesis inhibition (not ACCase)	thiocarbamate, chloro-carbonic-acid
O	action like indole acetic acid (synthetic auxins)	phenoxy-carboxylic-acid, benzoic acid
P	auxin transport inhibition	phthalamate, semicarbazone
Z	unknown	Pyrazolium, organoarsenical

Abbreviations: 4-HPPD: 4-hydroxy-phenyl-pyruvate-dioxygenase, ACCase: Acetyl CoA carboxylase, ALS: acetolactate synthase, AHAS: acetohydroxy-acid synthase, , DHP: dihydropteroate, PDS: phytoene desaturase step, PPO: protoporphyrinogen oxidase, VLCFAs: very long chain fatty acids

Inhibitors of the plant enzyme acetolactate synthase (EC 2.2.1.6, ALS), also known as acetohydroxy acid synthase (AHAS, EC 4.1.3.18), hinder catalysis of branched-chain amino acids valine, leucine and isoleucine.¹⁴¹ Disrupted branched-chain amino acid synthesis causes inhibition of DNA synthesis and ultimately cell death. Several classes of AHAS inhibitors, such as sulfonylureas and imidazolinones are known. Sulfonylurea-based herbicides are Amidosulfuron and Met-sulfuronmethyl. A prominent imidazolinone derivative herbicide is Imazapir, which was licenced in the USA in 1985. The first crystal structure of the catalytic subunit of yeast AHAS was released in 2002.¹⁴³ Since then, the mode of AHAS inhibition for herbicides was identified and more than ten crystal structures of different AHAS could be solved.¹⁴⁴ Some of these crystal structures include protein-bound herbicides. Based on co-crystallised herbicides, virtual screening and docking protocols have been used to identify novel AHAS sulfurea and imidazolinone-derived inhibitors.¹⁴⁵ 3D-QSAR was used to identify

new asymmetric aryl disulfides which showed *Arabidopsis* AHAS inhibition and herbicidal activity *in vivo*.¹⁴⁶ Molecular docking and 3D-QSAR were extended to DFT calculations for determination of HOMO and LUMO contribution of protein-ligand interactions of 32 isatin derivatives which proved to be active in both enzymatic assay and in *Brassica napus* root growth tests.¹⁴⁷

The herbicide Atrazine acts as an inhibitor of Photosystem (PSII). It is a triazine-derived herbicide. Atrazine is one of the most widely used herbicides, that can increase yield by 3-4%.¹⁴⁸ Its mode of action is the active site blockade of the plastoquinone-binding protein of PSII. This blockade causes the breakdown of electron transport processes that causes oxidative damage and plant death. Atrazine is toxic, having endocrine disruptive and possible carcinogenic effects.¹⁴⁸ It was also found that Atrazine reduces fish reproduction.¹⁴⁹ In 2004, Atrazine was banned from the European market due to groundwater pollution. In 2000, a computational modelling workflow was applied using homology modelling, docking and CoMFA for butenamide and quinone derivatives.¹⁵⁰ Another CoMFA study was published in which CoMFA models based on structural diverse classes of PSII inhibitors yielded useful knowledge for the development of novel PSII inhibitors which might be less toxic than Atrazine.¹⁵¹

1.6 Aim of this work

The main goal of this work is divided into two connected parts and is displayed in Figure 1.7. First, the development of a virtual screening workflow is desired which allows to search structural databases for potential *At*PARP inhibitors. The lack of structural data for the catalytic domain of *At*PARP proteins requires sequence analysis and homology modelling prior to derive a protein model of the target enzyme. Protein model evaluation will have to be performed to guarantee the model's applicability to perform the virtual screening process. This process is a multi-step procedure that consists generally of structure- and/or ligand-based pharmacophore filtering, followed by docking experiments that help to greatly reduce the numbers of structures in databases that will pass these filters. Validation of each filtering step will ensure that errors arising from both underlying mathematical and biological assumptions will influence the VS process in an appropriate and acceptable manner. After further application of structural database reducing *At*PARP-specific filters, potential candidates will be selected for more intensive investigation.

Meanwhile, at least one of the catalytic domains of *At*PARP should have been isolated and purified so far by cooperation partners and an enzyme inhibition assay should have been developed and validated. This will allow for testing selected candidate compounds for their inhibitory effects on the target and therefore affirm the aim of finding new *At*PARP inhibitors.

Once the structures, whose selection was based on the developed *in silico* workflow, have been verified as *At*PARP inhibitors, it should be investigated what is responsible for their inhibitory activity in a quantitative manner. This process is called quantitative structure-activity relationship (QSAR). QSAR can be used to gain insights into what chemical features might discriminate high-affine compounds from inactive or less active ones. The models also should be predictive so that activity of further VS outcomes could be estimated without the need of *in vitro* determination of inhibitory constants - depending on the quality of the QSAR model.

The second goal of this work is to integrate *At*PARP inhibition into the network of drought stress (DS) sensation and drought stress resistance. Parallel field trials that are designed to measure *At*PARP-verified inhibitors *in planta*, should be taken into consideration whether *in vitro* and *in planta* results can be compared and which practical significance *At*PARP inhibitors might have in the field. These results on the one hand might give further implications on the interplay of plant PARP and drought stress and on the other hand might be a starting point for the development of a new class of agrochemicals that increase crop yield under drought stress conditions.

Combined *in silico*, *in vitro* and *in planta* workflows

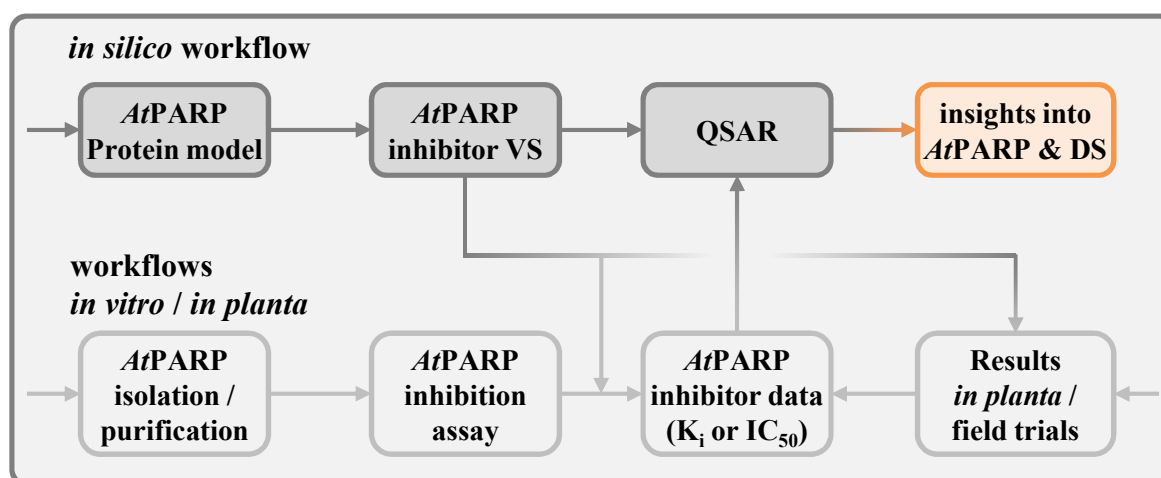


Figure 1.7: Workflow and aim of this work

2 Materials & Methods

2.1 Data sets

2.1.1 Natural substrate and natural substrate analogues

In PARP's catalytic reaction, the donor molecule nicotinamide adenine dinucleotide (NAD^+ , 11 in Figure 2.1) and a polymer of poly-ADP-ribose units - the acceptor molecule - are involved. The donor substrate NAD^+ was crystallised in Diphtheria Toxin (PDB entry 1TOX)¹²¹. The NAD^+ -analogue carba-NAD (CNA, 12 in Figure 2.1) was crystallised in the acceptor site of *Gg*PARP1 (PDB entry 1A26).³¹ In that experiment, the difference Fourier electron density map only allowed for the structural determination of the adenosine diphosphate moiety of carba-NAD (13 in Figure 2.1) because the nicotinamide moiety of carba-NAD was too mobile for structure determination. There is no crystal structure containing both the donor and acceptor structures. The co-crystallised ligands from 1TOX and 1A26 were used to investigate the positions of the substrates in the active site of *At*PARP1.

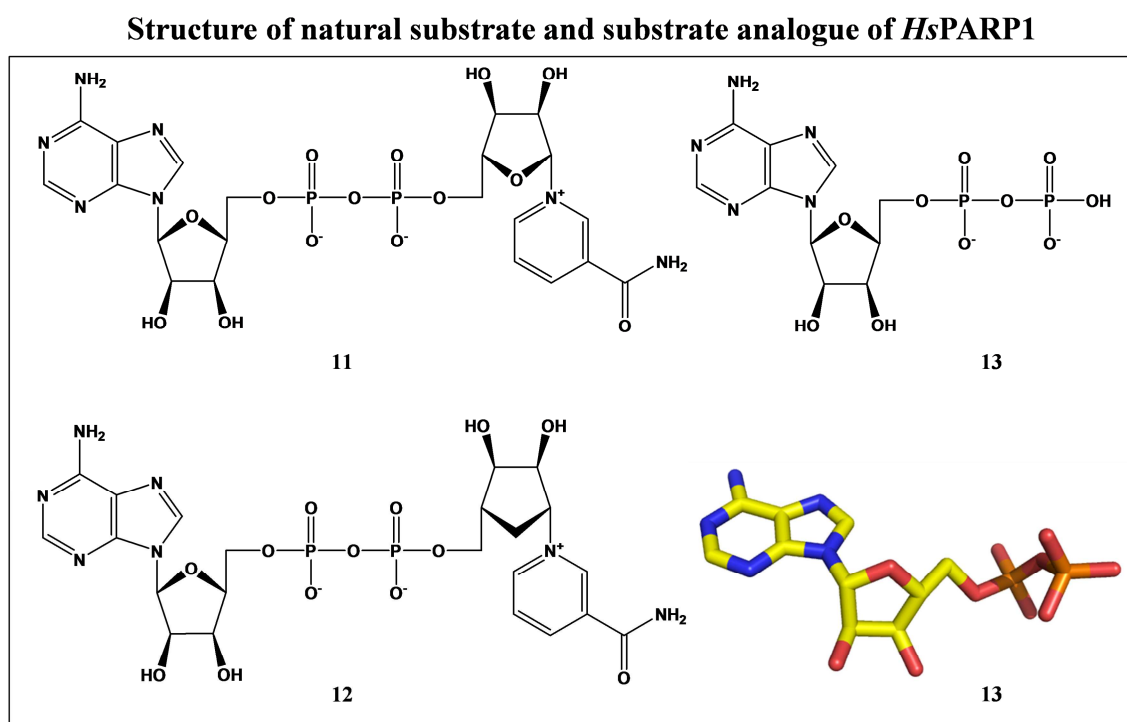


Figure 2.1: Natural substrate of *Hs*PARP1 and substrate analogue carba-NAD
Structures: 11 NAD^+ ; 12: carba-NAD; 13: moiety of carba-NAD for which electron densities were high enough for structure determination; 13 top: schematic representation of carba-NAD; 13 bottom: stick representation of carba-NAD in its conformation observed in *Gg*PARP1 (PDB entry 1A26)³¹

While carba-NAD represents the complete NAD^+ -mimicking structure, in the context of this

2.1.4 Known human PARP decoys

Decoy structures for *Hs*PARP1 were taken from the Directory of Useful Decoys (DUD).¹⁵⁸ They represent 1351 structures that are similar in scaffolds to human PARP inhibitors but do not bind to human PARP1. As for the human PARP1 inhibitors data set (2.1.3) it is assumed that they represent a random sample from the (unknown) population of human PARP1 decoys. Examples from the data set of *Hs*PARP1 decoys are shown in Figure 2.3

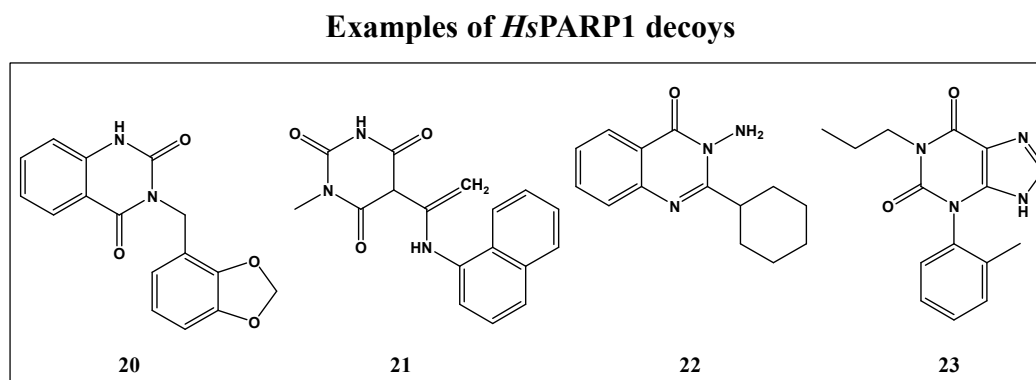


Figure 2.3: Examples of *Hs*PARP1 decoys

Structure names as given in the DUD.¹⁵⁸: 20: ZINC00424179, 21: ZINC02990370, 22: ZINC00818218, 23: ZINC00007652

2.2 Sequence analyses

2.2.1 Pairwise sequence alignment

Pairwise sequence alignments were performed using the program EMBOSS needle, which is available at the EMBL-European Bioinformatics Institute's (EMBL-EBI) webpage.¹⁵⁹ The program uses the alignment algorithm developed by Needleman and Wunsch¹⁶⁰ with the following EMBOSS needle's parameters: Matrix: BLOSUM62, gap penalty: 10.0, extend penalty: 0.5, gap penalty: false, end gap open: 10.0, end gap extend: 0.5

2.2.2 Multiple sequence alignment

Multiple sequence alignments were performed using the program Clustal Omega, available at the EMBL-EBI's webpage.^{161–163} The following parameters were used: KTUP: 1, Window: 5, Score: percent, Top Diagonals: 5, Pairgap: 3, Protein Weight Matrix: BLOSUM, Gap open: 10, Gap extend: 0.1, Protein Weight Matrix: BLOSUM, Gap open: 10, End gap: false, Gap extend: 0.2, Gap distance: 5, Iteration type: none, Number Iterations: 1, Clustering method: Neighbour-Joining algorithm.¹⁶⁴

2.3 Programs for Homology Modelling and Docking

2.3.1 Molecular Operating Environment (MOE)

In MOE, pharmacophores were created and databases searched for structures fulfilling the requirements of the pharmacophore. Molecular structures together with their chemical properties are stored in MOE in molecular database (*.mdb) files in which data manipulation, data processing and data analysis can be performed. The versions 2008.10¹⁶⁵, 2009.10¹⁶⁶, 2010.10¹⁶⁷, 2011.10¹⁶⁸ and 2012.10¹⁶⁹ of MOE were used.

The MOE docking Suite (MOE dock) was used for evaluation of docking performance on *Hs*PARP1 inhibitors (2.1.3) and decoys (2.1.4), as well as docking of NAD⁺ (Figure 2.1) into *Hs*PARP1. In the context of docking, the placement methods Alpha PMI, Alpha Triangle, Pharmacophore, Proxy Triangle and Triangle Matcher were used. Affinity dG scoring function was used to rate poses.

For conformational analysis of the *Hs*PARP1 decoy (2.1.4) and inhibitor (2.1.3) data sets, conformations of those structures were generated with the LowModeMD search¹⁷⁰ application with standard parameters, of which the most important are the rejection limit = 100, RMS gradient = 0.005, RMSD limit = 0.25, Energy Window = 7 kcal/mol.

2.3.2 POSIT

Open Eye's software application POSIT (version 1.0.0) was used for identification of bioactive poses of compounds that are known to bind to *At*PARP. POSIT compares ligand poses (e.g. generated by a docking program) to X-ray crystal coordinates and calculates a probability that a generated pose is correct. POSIT uses measures of similarity to define a probability. These measures are shape comparisons such as the 3D *Tanimoto*_{Combo}¹⁷¹ and 2D path -based fingerprints, as well as the Mills Dean approximation of electrostatics.¹⁷² The value of the *Tanimoto*_{Combo} is the sum of *Tanimoto*_{shape} and *Tanimoto*_{Color}, where the *Tanimoto*_{shape} of two structures *A* and *B* is:

$$\begin{aligned} & \mathbf{Tanimoto}_{shape}(A, B) \\ &= \frac{\int \mathbf{A}(\vec{r}) * \mathbf{B}(\vec{r})}{\int \mathbf{A}(\vec{r}) * \mathbf{A}(\vec{r}) + \int \mathbf{B}(\vec{r}) * \mathbf{B}(\vec{r}) - \int \mathbf{A}(\vec{r}) * \mathbf{B}(\vec{r})} \end{aligned} \quad (2.1)$$

$Tanimoto_{shape}$ can take values from 0 for non-overlapping voxels of two structures to 1 for two structures sharing the same voxels. Thereby, a voxel represents the volume of a cube of structures. Each voxel is assigned a colour representing chemical features like hydrogen bond

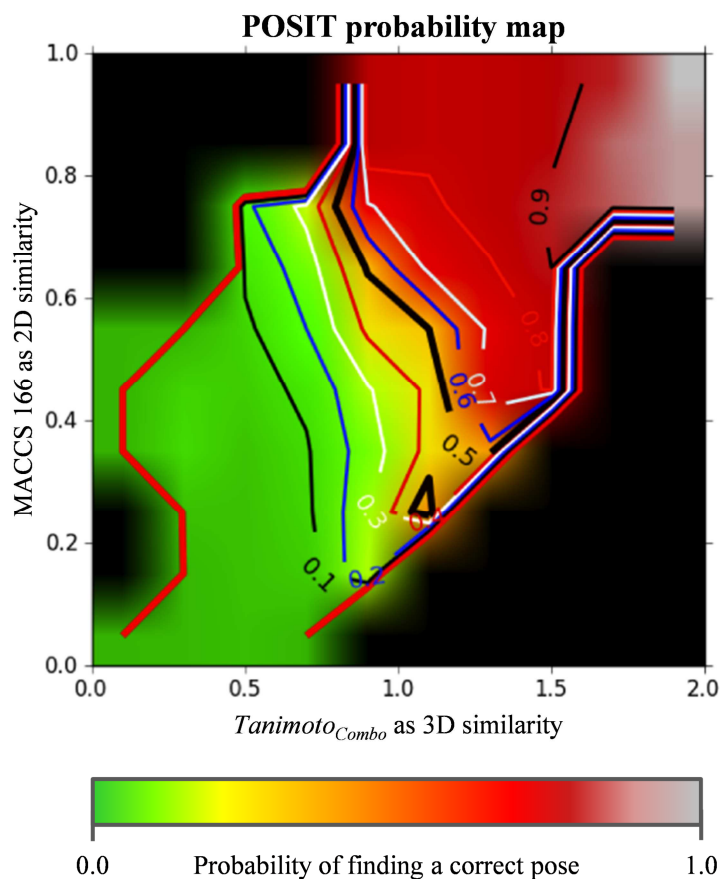


Figure 2.4: POSIT probability map, modified from POSIT manual

method.¹⁷³ As a next step, ligand protein optimisation is performed to remove steric clashes and improve interactions between the ligand and the protein. Finally, a probability is provided, which, given the ligand does bind to the protein, is the likelihood of the POSIT pose being the one, that one would observe in a crystal structure. This probability is a function of the $Tanimoto_{Combo}$ and the MACCS166 key fingerprint and is displayed in Figure 2.4.

2.3.3 YASARA

YASARA¹⁷⁴ was used for Homology Modelling of *At*PARP, energy minimisation of protein models and Molecular Dynamic simulations.¹⁷⁵ For energy minimisation of proteins, the YASARA2 force field was used.¹⁷⁴ The AMBER03 force field¹⁷⁶ was used for Molecular Dynamic simulations which are further described in 2.3.3.2. Also, evaluation of MD simulations was performed in YASARA.

donors. For $Tanimoto_{color}$, colour-coded voxels are used for calculation. It also takes values from 0 to 1, where 1 equals the same chemical features of structures *A* and *B*. Therefore, the maximum value, that $Tanimoto_{Combo}$ can take on, is 2.

Among a set of X-ray complexes, POSIT chooses the one whose bound ligand matches a predicted pose (of a new) ligand best by applying the similarity measures. For the chosen complex, to better match the binding mode of the bound ligand, a flexible fit is performed using an adiabatic optimisation

2.3.3.1 Homology modelling in YASARA

The first step in homology modelling is to select a protein template. This is chosen from the PDB in general.⁶⁹ The three-dimensional structure of the target, based on its primary sequence, can then be modelled using YASARA's macro `md_build.mcr`. The standard parameters are: PSI-BLAST iterations = 6, PSI-BLAST E-value = 0.5, oligomerization state = 4, used templates = 5, Alignments per template = 5, terminal extension = 10, loop samples = 50, modelling speed = slow. Whenever homology modelling with YASARA was used in this work, a specific template was selected ahead and according to this, the parameter "used templates" was altered from 5 to 1. The homology modelling workflow in YASARA consists of the following stages (Table 2.1) if a single template is chosen:

Table 2.1: Homology modelling steps in YASARA

Stage no.	Sequence length
1	Setting modelling parameters
2	Perform a BLAST search to retrieve (multiple) sequence alignment
3	Predict the secondary structure using the Discrimination of Secondary structure Class (DSC) prediction algorithm ¹⁷⁷ and loop refinement
4	Creation of tertiary structure of the query sequence with subsequent loop modelling
5	Side chain optimisation followed by a combined steepest descent and simulated annealing minimisation with fixed backbone atoms
6	Full unrestrained simulated annealing for refinement
7	Model evaluation using a Z-score defined as $\mathbf{Z} = \mathbf{0.145} \times \text{dihedrals} + \mathbf{0.390} \times \text{packing1D} + \mathbf{0.465} \times \text{packing3D} \quad (2.2)$
8a	If more than one solution was built for sequence alignment, steps 1-7 are performed for all remaining alignment solutions, A final hybrid model, based on all previous homology models is created and evaluated.
8b	Finally, among all models, the one with the highest positive Z -score is chosen as the final homology model of the query sequence

2.3.3.2 MD simulations in YASARA

Molecular Dynamics (MD) simulations were performed using the tool `md_run.mcr` in YASARA (version 12.11.25) with the AMBER03 force field.¹⁷⁶ Intermolecular and intramolecular forces were calculated every 1 fs resulting in a simulation time step of 1 fs. After a simulation time of 5.000 simulation time steps (5.000 fs), a snapshot of the current simulation system was saved. In total, 4.000 simulation snapshots were saved, resulting in a total simulation time of 20 ns.

To set up the simulation, YASARA's Neutralization Experiment was used to predict pK_a values using Ewald summation.¹⁷⁸ The simulation cell was defined being 10 Å greater than the protein in each direction. It was simulated at pH = 7.0, with a physiological NaCl concentration of 0.9% and a density of water of 0.997 g/l. Temperature was controlled by rescaling atom velocities using a Berendsen thermostat¹⁷⁹ based on the time-averaged temperature.¹⁷⁵ It was simulated using periodic boundary conditions and long-range Coulomb interactions were calculated using Particle Mesh Ewald (PME) algorithm.¹⁸⁰

All MD simulations were conducted as independent triplicates as are required in the Journal of Molecular Modelling. Since in YASARA, initial kinetic energies and atom velocities are assigned randomly but fixed for a given temperature, independent triplicates were generated by running each simulation at slightly different temperatures of 298 ± 0.0001 K. Therefore, by selecting different, initial atom velocities were set in an independent manner while merely affecting the average kinetic energy of the system.

2.3.4 ConfGen

ConfGen^{181,182}, a tool of Schrödinger Software, was used to generate bioactive conformers of structures. It was used with the intermediate search strategy and the following settings: The maximum number of search steps is 1000, the number of conformers generated per rotatable bond is 75, an RMSD value of 1.0 Å is used to detect redundant conformers. All conformers having an energy more than 25 kcal/mol (104.67 kJ/mol) higher than the lowest energy conformer are eliminated. The minimum dihedral angle difference for polar hydrogens is 60°. The maximum relative energy for flexible rings is 2.39 kcal/mol (10 kJ/mol) and the energy threshold for periodic torsions is 5.74 kcal/mol (24 kJ/mol). The total number of ring conformations per ligand is 16, the number of ring conformations for a single ring is 8. All remaining conformers were energy minimised using the OPLS 2005 force field.¹⁸³ In this study, ConfGen version 2.2 from Schrödinger Suite 2010, was used.

2.3.5 LigPrep

LigPrep¹⁸⁴ by Schrödinger Software was used to generate low energy 3D output structures being variations of ionisation state and tautomers of the input structures. Ionisation states were generated at target pH of 7.0 ± 2.0 . Possible protonation states were generated using the program Epik¹⁸⁵⁻¹⁸⁷. Chirality was changed only if it was not specified before. All other stereo

centres were retained. At most 32 stereoisomers per input structure were generated. LigPrep version 2.4, as implemented in Schrödinger Suite 2010, was used in this study.

2.3.6 Glide

Glide^{188,189} (Grid-based Ligand Docking with Energetics) is the docking program by Schrödinger Software.¹⁹⁰ The docking process is divided into four stages. During the first, a site-point search is performed. If the first stage is passed, stage two begins with evaluation of steric clashes (Diameter test), followed by a subset test in which hydrogen bonding and ligand-metal interactions are taken into account and scored. If this score passes a threshold, all interactions are considered and scored, too (greedy score). Scoring in these stages is done using ChemScore scoring function.¹⁹¹ The third docking stage contains an energy minimisation using pre-computed OPLS-AA grids for the receptor. The final step consists of scoring the remaining poses with Schrödinger GlideScore (GScore), a modified version of the ChemScore¹⁹¹ scoring function:

$$\text{GScore} = 0.065 \times \text{vdW} + 0.130 \times \text{Coul} + \text{Lipo} + \text{HBond} + \text{Metal} + \text{BuryP} + \text{RotB} + \text{Site} \quad (2.3)$$

Glide Extra Precision Mode (Glide XP)¹⁹² uses a more exhaustive sampling strategy than the standard Glide (standard precision, SP) docking protocol, but also a modified scoring function compared to Glide SP. It is designed to recognise false positive ligands by recognising poses of ligands which are unfavourable and then removing them.

The receptor for Glide docking was prepared using the Receptor Grid generation tool. In this study, version 5.6 of Glide as part of Schrödinger Suite 2010, was used.

2.3.7 GOLD

The docking program GOLD (Genetic Optimization for Ligand Docking, version 5.0.1) was developed in 1995.^{193,194} It uses a genetic algorithm (GA)¹⁹⁵ that enables to rapidly explore the conformational flexibility of a ligand and sampling binding modes into a binding site that is treated partially flexible. Genetic algorithms are in principle able to find an optimum solution to optimisation problems which makes it an interesting application in performing conformational analysis of small and flexible molecules.^{194,196,197} The scoring functions ASP¹⁹⁸, ChemScore^{191,199}, GOLDScore^{193,200} and ChemPLP²⁰¹ were used.

2.3.8 PLANTS

The docking program PLANTS (Protein Ligand ANT System, version 1.1) uses ant colony optimisation (ACO) which, together with particle swarm optimisation (PSO) methods, constitutes main swarm-intelligence approaches. These approaches belong to the class of stochastic optimisation methods that can be used to find a global minimum structure with respect to a given objective function f .

$$\min_{\bar{x} \in \mathbb{R}^n} f(\bar{x}): \mathbb{R}^n \rightarrow \mathbb{R} \quad (2.4)$$

In the context of docking the objective function f is called scoring function. Here, $\bar{x} = [x_1, \dots, x_n]^t \in \mathbb{R}^n$ represents the protein's (n_p) and ligand's (n_l) degrees of freedom with total degrees of freedom $n = 6 + n_l + n_p$. PLANTS uses two empirical scoring functions $PLANTS_{PLP}$, $PLANTS_{ChemPLP}$, that are derived to reproduce experimentally determined protein-ligand complexes. $PLANTS_{PLP}$ uses a distance-based piecewise linear potential and is adapted from the work of Gelhaar²⁰² and Verkhivker²⁰³. The second scoring function, the one, which was used for the docking studies used in this context, is $PLANTS_{ChemPLP}$. It is of the form:

$$f_{PLANTS_{CHEMPLP}} = f_{PLP} + f_{chem-hb} + f_{Ltors} + f_{Lclash} + 0.3 * f_{Protscore} - 20 \quad (2.5)$$

Steric interactions between the protein and the ligand are calculated by f_{PLP} . The second term, $f_{chem-hb}$, describes hydrogen bonding and metal-acceptor interactions between the protein and the ligand. Weak CH-O interactions are considered by differentiating charged and neutral hydrogen bonds as has been done by Verdonk.²⁰⁴ It is adapted from the ChemScore scoring function, as it is implemented in GOLD.²⁰⁵ Intramolecular ligand scoring terms consist of a clash term f_{Lclash} and a torsional potential f_{Ltors} , adapted from Clark and co-workers.²⁰⁶ The same potential as f_{PLP} describes the Intramolecular protein-interactions ($f_{Protscore}$), together with an intra-side chain clash term.

PLANTS allows the weights for hydrogen bond contributions w_{hb} to be changed. In standard scoring parameters, the hydrogen bond weights, w_{hb} , are set to $w_{hb} = 1$. In order to direct the docking procedure and reward specific protein-ligand interactions, these parameters have been changed, e.g. increased to $w_{hb} = 10$, which results in poses with better docking scores in which hydrogen bond that have increased weights, exist (2.6.4).

2.4 Application-dependent Homology Modelling

The large amount and variety of PARP and ADPRT-like crystal structures in the PDB enables for the usage of specific homology modelling techniques as well as using specific templates for homology modelling of *At*PARP. Depending on the objective (investigation of homology model stability, the positions of PARP's natural substrates and the bioactive conformation of new *At*PARP inhibitors), it was taken advantage of the range of potential templates and homology modelling techniques as will be described in the following sections.

2.4.1 Investigation of protein stability

The information of from two X-ray structures of the catalytic domain of *Gg*PARP1 was used for investigation of protein stability. PDB entry 2PAX contains *Gg*PARP1₆₅₉₋₁₀₀₈ and the PARP inhibitor 4-amino-benzo[de]isoquinoline-1,3-dione (PDB identifier 4AN, 16, Figure 2.2) in the active site.⁷² Using 4AN as reference ligand for *At*PARP1 *in silico* experiments was advantageous because this compound was recently shown to be a new *At*PARP1 inhibitor.^{207,208} PDB entry 2PAW contains *Gg*PARP1₆₅₉₋₁₀₀₆ without an inhibitor bound in the active site.⁷² After building homology models of 4AN-ligated and unligated *At*PARP1 models, structural changes during an MD simulation of the models were investigated by analysing the root mean squared fluctuations (RMSF) (2.7.8) for all C_α atoms. MD simulations were carried out in YASARA under specifications described in 2.3.3.2. The RMSF were compared to C_α B-factor distributions of the X-ray structures in the presence and absence of the inhibitor. The workflow is shown in Figure 2.5. By investigating the B-factor distributions for ligated and unligated *Gg*PARP1, two hypotheses were tested:

The first hypothesis was: The MD simulation-based RMSF distributions for the ligated and unligated *At*PARP1 models should be of same shape as the experimentally derived B-factor distributions for ligated and unligated *Gg*PARP1 crystal structures.

The second hypothesis was based on the observation, that upon binding of 4AN in *Gg*PARP1 a stabilising effect occurs in the region of the loop in proximity to Tyr₉₀₇ (corresponding to Tyr₉₀₇ in *Hs*PARP1 and Tyr₅₃₁ in *At*PARP1).⁷² Therefore, the second hypothesis was: If Tyr₅₃₁ in *At*PARP1 has the same stabilising effect upon inhibitor binding as it was observed in *Gg*PARP1, the same effect should be detected as decreased RMSF values for 4AN-ligated *At*PARP1 during MD simulations (2.3.3.2) in the region around Tyr₅₃₁ in comparison to unligated *At*PARP1 RMSF values in the same region.

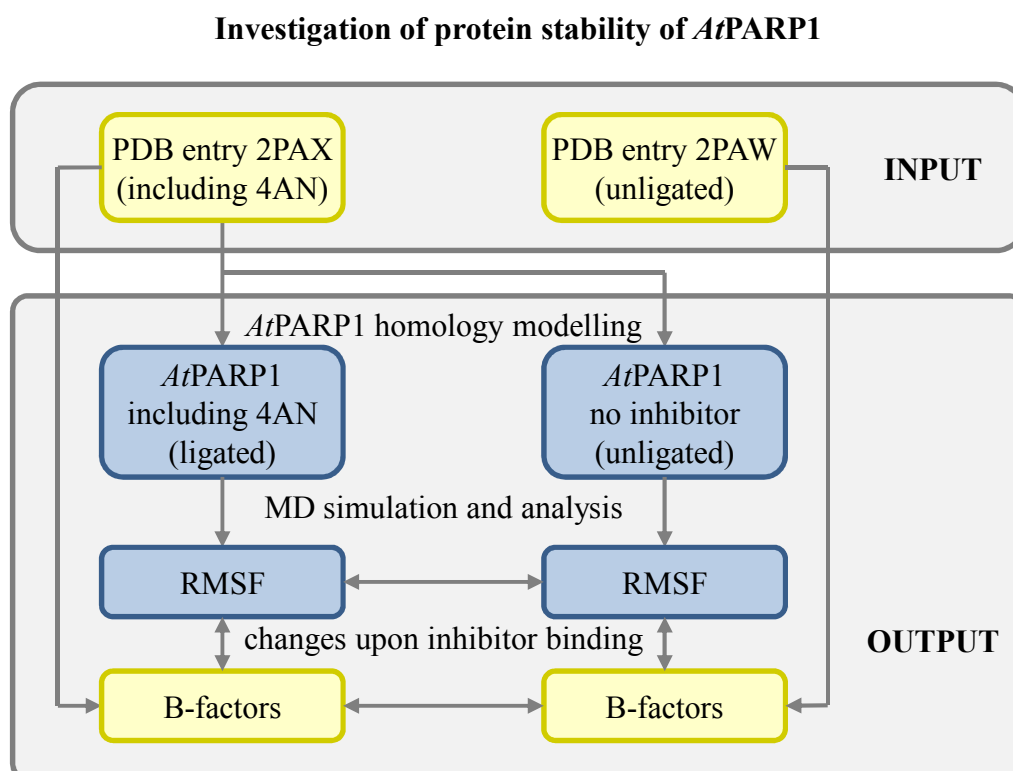


Figure 2.5: Workflow of investigation of protein stability in *At*PARP1

2.4.2 Investigation of positions of natural substrates

The structures of NAD⁺ and CNA (Figure 2.1) were used for investigation of *At*PARP1's natural substrate positions. Their positions in the active site of *At*PARP1 were investigated by an approach that uses superpositioning of functional conserved active sites and that is making use of conserved residues in ADP-ribosylating toxins and PARP enzymes. A schematic representation of the workflow is shown in Figure 2.6.

For the superpositioning of natural substrate in *At*PARP1, the ProBiS web server²⁰⁹ was used, which identifies structurally similar binding sites. The web server uses the ProBiS algorithm²¹⁰ which compares a query protein structure with the non-redundant PDB (nr-PDB) database which (since 30th Nov. 2013) contains 37.643 entries. In the algorithm, a query protein is compared to each entry, e.g. each protein, in the ProBiS database²¹¹. For each protein-protein comparison, the ProBiS algorithm represents a protein as a three-dimensional graph of vertices and edges, where a single vertex represents a physicochemical property of a functional group of a surface-accessible amino acid. Physicochemical properties are divided into hydrogen bond acceptor (AC), hydrogen bond donor (DO), mixed acceptor/donor (ACDO), aromatic (PI) and aliphatic (AL) (shown in Figure 2.7), according to the physicochemical properties proposed by Schmitt and co-workers.²¹²

Investigation of positions of *At*PARP1's natural substrates – ProBiS approach

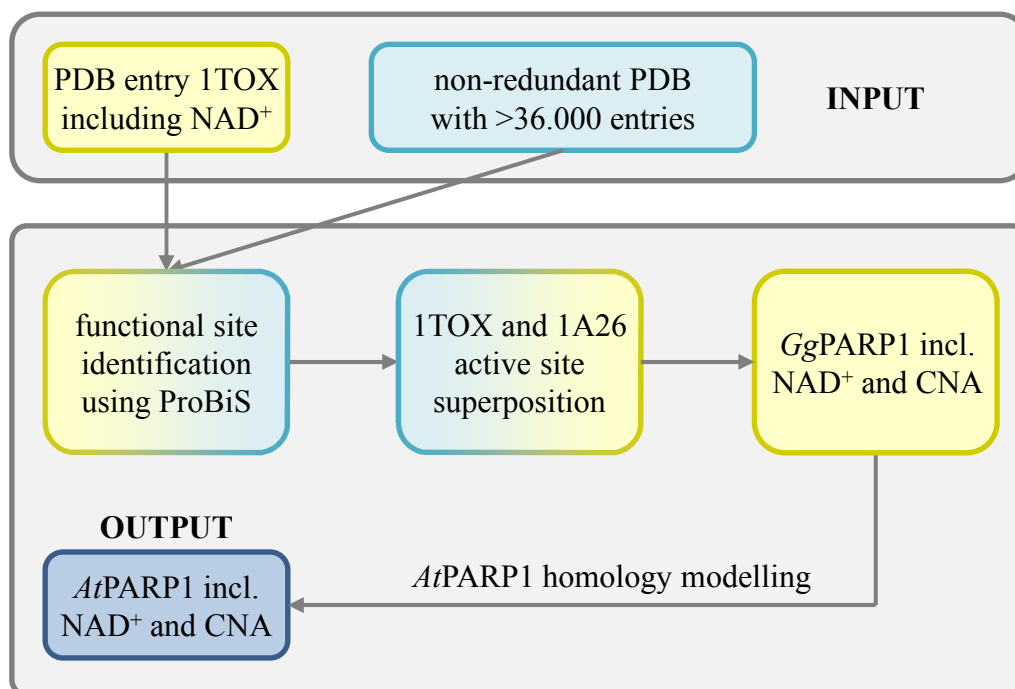


Figure 2.6: Workflow using ProBiS

From these 3D graphs, subgraphs are generated which are defined as all vertices in a radius smaller 15 Å radius from a central vertex. Similar subgraphs between the query and database

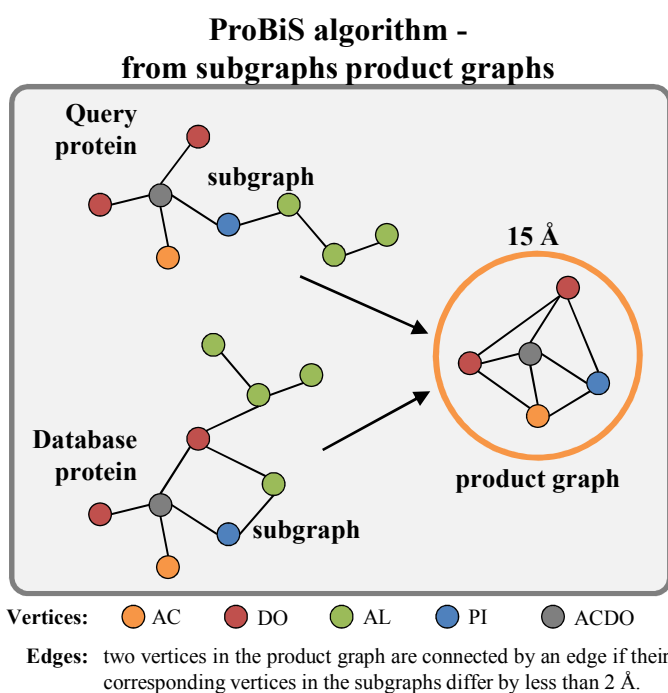


Figure 2.7: Schematic representation of ProBiS algorithm I Product graph generation of subgraph of query and database proteins in ProBiS algorithm

protein are found by calculating a similarity value between the subgraphs.²¹³ If two subgraphs are defined as similar, a product graph is constructed, as depicted in Figure 2.7.

In a next step, a maximum clique algorithm²¹⁴ is applied to find a maximum clique in all product graphs, which corresponds to common substructure consisting of the maximum number of vertices. Each maximum clique can be regarded as a local structural alignment of two proteins. The

statistical significance of each structural alignment is assessed by a surface vector angle (which has to be smaller than 90°), an RMSD value (which has to be smaller than 2 \AA) and an E -value (using the Karlin-Altschul equation²¹⁵⁻²¹⁷, which has to be below the threshold of $1.0 \cdot 10^{-4}$).

Statistically significant maximum cliques with more than 5 vertices are clustered. Statistically significant local structural alignments in the ProBiS Database are calculated using a Z-score, which is derived from a standardised alignment score, the RMSD of two pairs of superimposed vertices and the calculated E -value. A schematic representation of the ProBiS algorithm is shown in Figure 2.8.

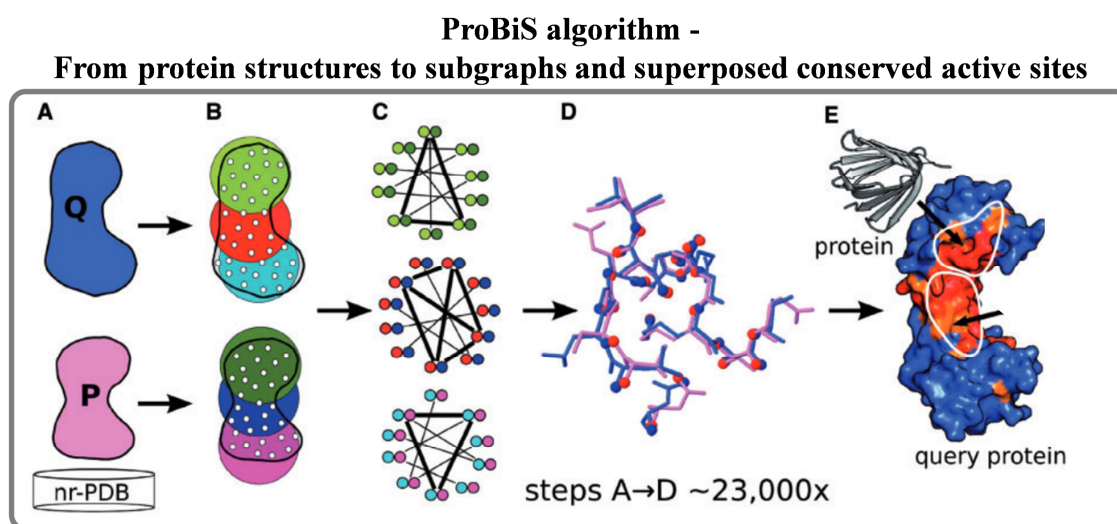


Figure 2.8: Schematic representation of ProBiS algorithm II

Scheme modified from Konc, J. & Janežič²¹⁰ A: Comparison of a query protein with all proteins contained in a database B: Representation of the query and database protein as graphs (and subgraphs) consisting of vertices and edges C: Generation of product graphs and maximum cliques within product graphs D: Production of structural local alignments as results of each maximum clique. E: Repetition of preceding steps for all proteins in the database

As a result, the ProBiS algorithm results a list of proteins structurally similar to the query protein. For each entry of the list, the superposed query and database protein structures, as well as all corresponding significance measures, can be downloaded as .pdb files for further usage.

2.4.3 Investigation of bioactive conformation of *At*PARP inhibitors

Investigations of bioactive conformations of new putative *At*PARP inhibitors were performed with OpenEye's POSIT (2.3.2). POSIT finds a ligand's probable bioactive binding pose by comparing conformations of a confirmed bioactive ligand to known binding poses of other

ligands for the same target. Based on these comparisons, it assigns a new binding pose for the new ligand structure that is most similar to the binding pose of the known ones. This process is solely executable for identical targets. Therefore, to predict bioactive conformations for new *At*PARP inhibitors (for which no crystal structures are known so far), crystal structures of PARP orthologues (eg. *Hs*PARP, *Gg*PARP1) with inhibitors bound served as indirect templates of known binding poses. For each of the inhibitor-complexed PARP orthologues, the structure of *At*PARP1 was homology-modeled as a first step.

Among all X-ray structures representing catalytic domains of PARP enzymes, 18 X-ray structures from *Hs*PARP1, *Hs*PARP2, *Hs*PARP3, *Gg*PARP1 were used to generate homology models of *At*PARP1 using YASARA (2.3.3.1). Following homology modelling, each *At*PARP1 model contained the corresponding template inhibitor which, for the purpose of this approach, was assumed to be inhibiting *At*PARP1, too. The models were prepared with OpenEye's combine_receptor tool. For each of the confirmed *At*PARP inhibitors (3.8) it was calculated, which of the 18 *At*PARP1 models contains an inhibitor conformation that is most similar to conformations of a new inhibitor. The *Tanimoto*_{Combo} score (2.3.2) was taken as a measure and a probability was generated. The workflow of this bioactive conformation research is shown in Figure 2.9.

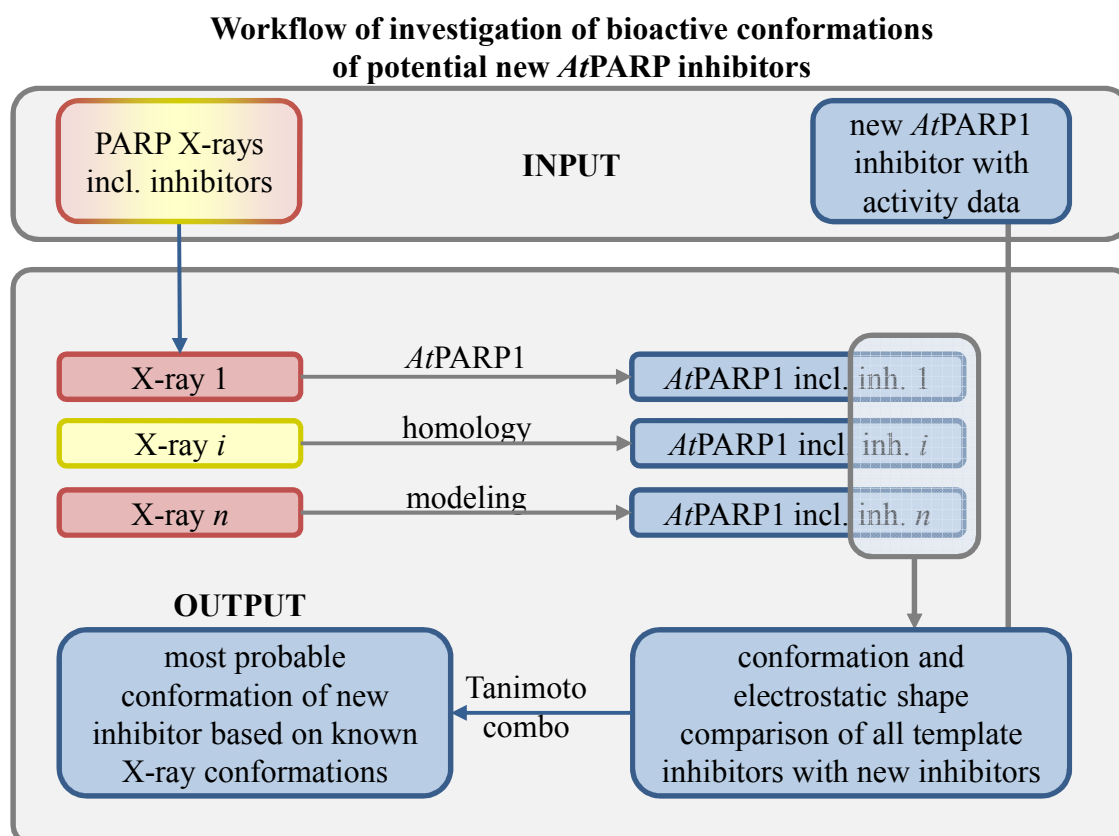


Figure 2.9: Finding most probable bioactive conformations of *At*PARP inhibitors

2.5 Pharmacophore creation

To decide whether a conformation of a structure fits into the active site of *At*PARP2, a pharmacophore was built using MOE's pharmacophore query editor tool with the unified annotation scheme.¹⁶⁸ The Pharmacophore was derived from the *Hs*PARP1 nicotinamide-mimicking pharmacophore in which two hydrogen bonds between the conserved Gly₈₆₃ and the inhibitor, and hydrophobic contacts between Tyr₉₀₇ and an electron-rich aromatic ring-system need to be present for inhibitor recognition and binding. The pharmacophore was created based on the *At*PARP2 homology model (3.2). It consists of the required hydrogen bond donor and acceptor spheres, as well as the hydrophobic interaction area of the inhibitor defined by the hydrophobic interaction centre and two interaction vectors orthogonal to the plane of the NA-mimicking aromatic ring of the modelled ligand FRQ (Figure 2.10, structure 24, from PDB entry 1UK1¹⁵⁷). An excluding volume shell was defined for all atoms having a distance greater than 4.5 Å around the modelled inhibitor FRQ to define the shape of the active site. All pharmacophore-defining entities were defined as being essential (5.3). The schematic representations of the PARP pharmacophore, as well as the created *At*PARP2 pharmacophore in MOE, are shown in Figure 2.10.

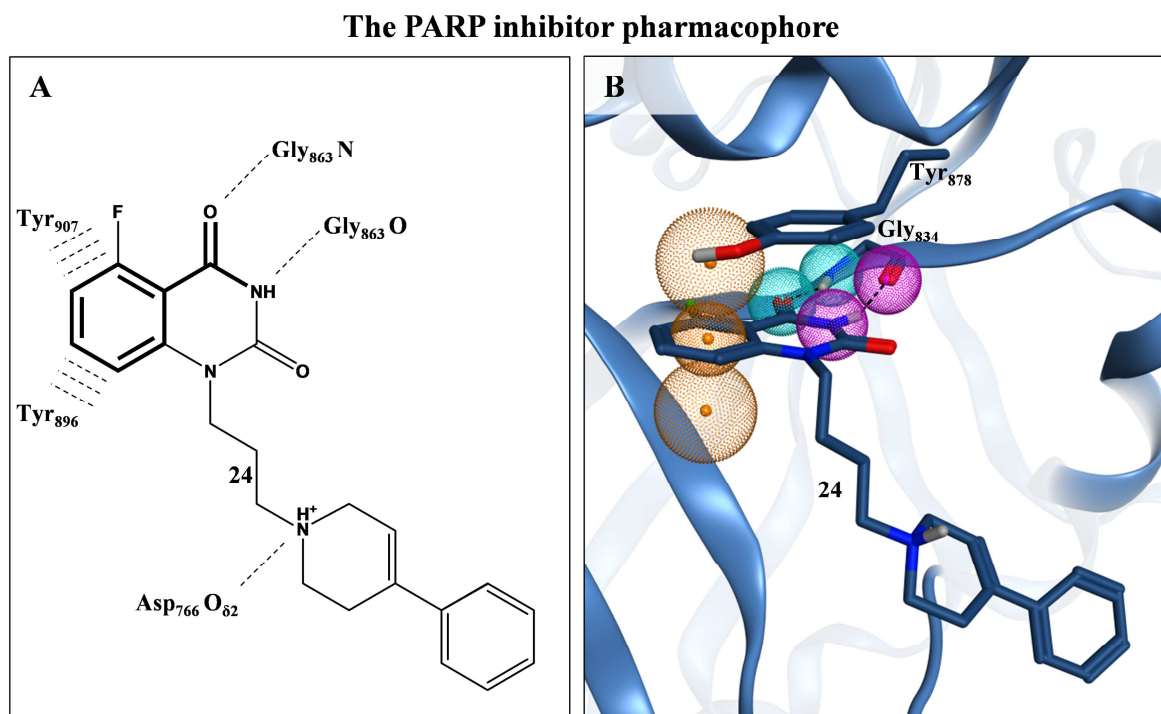


Figure 2.10: The PARP pharmacophore

A: Representation of the PARP pharmacophore (*Hs*PARP1 numbering) B: view into the active site of *At*PARP2 with pharmacophore spheres and FRQ being present; 24: *Hs*PARP1 inhibitor FRQ, co-crystallised in *Hs*PARP1 (PDB entry 1UK1)¹⁵⁷; excluded volumes not shown for clarity reasons

2.6 Docking procedure

2.6.1 General aspects

Establishing a docking procedure and assessing the quality of a docking procedure for *At*PARP is not straightforward because there is not much knowledge available. First, only NA and 3AB have been verified as *At*PARP inhibitors so far²¹⁸, and 4AN was only recently shown to inhibit *At*PARP1, too.^{207,208} Furthermore, there are no decoy structures for any plant PARP known to this point. In addition to that, there is no X-ray structure of any *At*PARP's catalytic domain including a co-crystallised inhibitor in its active site deposited in the PDB. These facts would be preconditions to apply a docking procedure or define a docking threshold to discriminate true inhibitors from decoys in a direct way. In contrast to *At*PARP, these requirements are fulfilled for *Hs*PARP1. To make use of this knowledge and establish a docking threshold for *At*PARP inhibitors, the following steps as listed in Table 2.2 were performed:

Table 2.2: Steps to be performed to define an *At*PARP docking procedure

Step	Task
Definition of a docking threshold for human PARP1	
1.1	definition of data sets for <i>Hs</i> PARP1 ligands and <i>Hs</i> PARP1 decoys
1.2	choosing the most suitable docking program for this purpose
1.3	establishing a docking procedure for selected data sets and verification
1.4	definition of criteria for discrimination of decoy and ligand structures and derivation of a docking threshold from these criteria
1.5	investigation of docking performance under these conditions
Performing <i>Hs</i> PARP1 docking procedure on <i>At</i> PARP1 and <i>At</i> PARP2	
2.1	defining molecular, biological and statistical assumptions under which the <i>Hs</i> PARP1 docking procedure can be transferred onto the <i>At</i> PARP1/2 docking procedure
2.2	application of docking procedure, that was established for <i>Hs</i> PARP1, by using analogue conditions as for the <i>Hs</i> PARP1 procedure and incorporate underlying assumptions
Definition of a new docking threshold for <i>At</i> PARP1/2	
3.1	investigation of differences of docking procedure of <i>Hs</i> PARP1 and <i>At</i> PARP1/2
3.2	derivation of new docking threshold for <i>At</i> PARP1/2

Characteristics derived from the *Hs*PARP1 docking procedure can be analysed after compounds from that database are bought and tested on *At*PARP1 in validated *in vitro* assays. The docking workflow starting from human PARP and resulting in selection of potential *At*PARP inhibitors is shown in Figure 2.11

Workflow for establishing an *Hs*PARP1-derived *At*PARP docking procedure

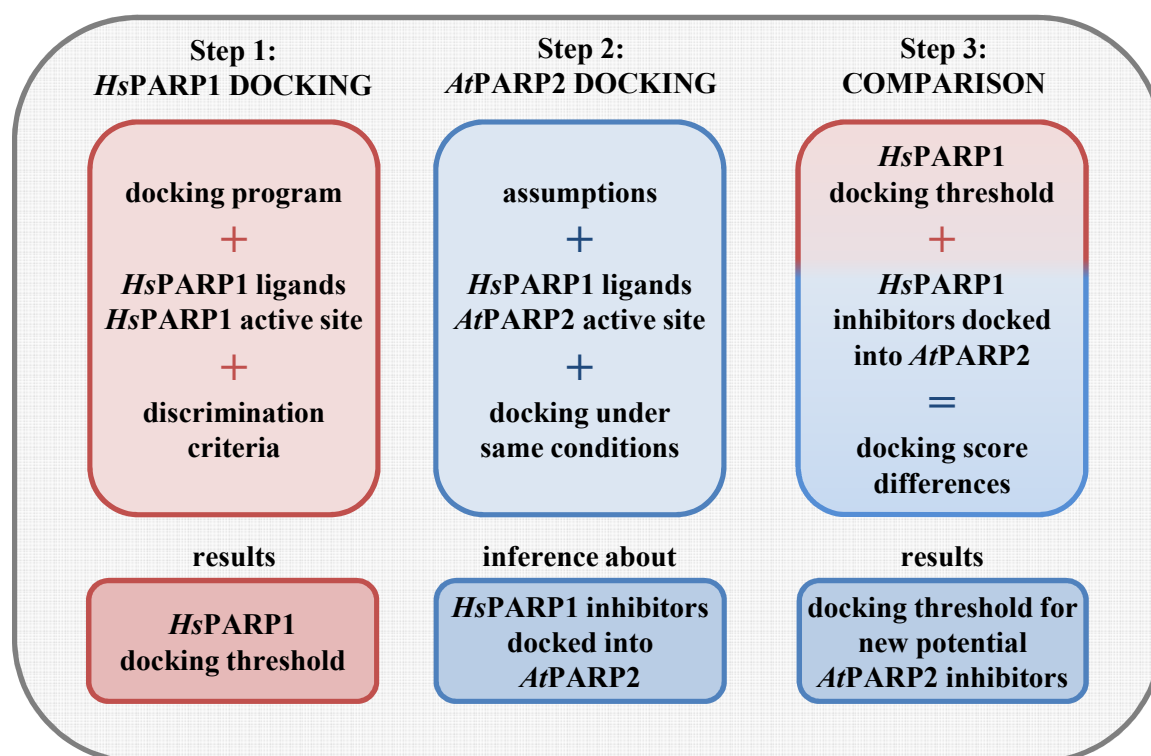


Figure 2.11: Docking workflow for establishing an *At*PARP docking procedure

Step 1 identifies a docking threshold for *Hs*PARP1, Step 2 applies all steps performed in steps 1 onto *At*PARP1 and Step 3 transfers this into a development of a new threshold for *At*PARP

2.6.2 Data sets

Two data sets for establishing a docking threshold were used. First, the Novikov data set described in 2.1.3 that contains a sample of *Hs*PARP1 inhibitors ($n = 142$) and secondly, the decoy data set described in 2.1.4 which contains known human PARP structures which are known not to bind to *Hs*PARP1 ($n=1351$).

2.6.3 Docking programs

The docking suite implemented in MOE, MOE dock, was used with five different placement routines: Alpha PMI, Alpha Triangle, Pharmacophore, Triangle Matcher and Proxy Triangle. For each placement routine, three refinement strategies for docking poses were used. First, no refinement at all was performed for direct placement. Second, tethered refinement of all non-hydrogen side chain atoms with tethering factor 10 was performed allowing partial refinement of the active site during the ligand's placement. As a third strategy during ligand placement, the active site's amino acid side chain atoms were set free without any tethering. This allowed for more complex conformational changes during placement of a ligand in the active side.

These settings result in 15 different docking routine combinations. In each routine, the reference ligand FRQ (24) was defined as the centre of the active site. Affinity dG was used as scoring function.

The scoring function extra precision glide (Glide XP) was used in the docking program Glide.^{188,189,192} The four scoring functions ASP, ChemPLP²⁰¹, ChemScore^{191,199} and GOLDScore^{193,200} were used in GOLD²⁰⁰. Scoring functions PLP, PLP95 and ChemPLP were used in PLANTS.^{201,219}

The binding site in which a ligand is placed during docking is defined differently in all docking programs. To define the binding site as similar as possible for all docking programs, the following settings have been used: FRQ was used as reference inhibitor in MOE (2.3.1), GOLD (2.3.7) and Glide (2.3.6). In PLANTS (2.3.8), the centre of the atomic coordinates of FRQ and a surrounding shell of 12 Å around this centre defined the active site. In GOLD and PLANTS, amino acid side chains, that participate in the known PARP pharmacophore, being Tyr₉₀₇/ Tyr₅₃₁/ Tyr₈₇₈ (*Hs*PARP1/ *At*PARP1/ *At*PARP2 numbering) respectively and Ser₉₀₄/ Ser₅₂₈/ Ser₈₇₅ respectively, were defined as flexible. Also, upon inspection of the active sites, Glu₇₆₃/ Glu₃₈₈/ Lys₇₃₅, respectively, were defined as flexible. The flexibility of side chains in GOLD was defined by not using rotamer libraries but by allowing full rotation about rotatable side chain bonds.

2.6.4 PARP pharmacophore-directed docking

To improve the identification of correct poses, the docking protocols have been adjusted. In Glide (2.3.6), GOLD (2.3.7), and PLANTS (2.3.8), the weights, w_i , for rating hydrogen bonds between the protein and the inhibitor were changed from the default value of 1.

In *Hs*PARP1, there is a hydrogen bond between the backbone nitrogen of Gly₈₆₃ and the

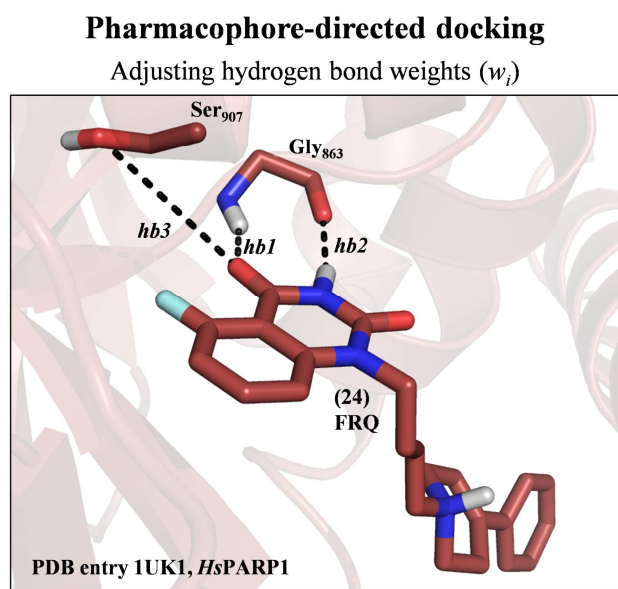


Figure 2.12: Hydrogen bond weights adjusted for pharmacophore-directed docking
View into active site of *Hs*PARP1 showing inhibitor FRQ being hydrogen bonded to Gly₈₆₃ and Ser₉₀₄.

inhibitor, *hb1*, and another hydrogen bond between the carbonyl oxygen atom of Gly₈₆₃ and the inhibitor, *hb2*. Both hydrogen both weights, w_{hb1} and w_{hb2} , were increased. Furthermore, the weight of the hydrogen bond between the side chain atom O_γ of Ser₉₀₄ and the inhibitor, *hb3*, was increased 10-fold (such that $w_{hb1} = w_{hb2} = w_{hb3} = 10$). In MOE, the hydrogen bond interactions *hb1* and *hb2* between the Gly₈₆₃ and the inhibitor were modelled by incorporation of the pharmacophore features described in 2.5.

A view into the active site of *Hs*PRAP1,

including its inhibitor FRQ (24) and the corresponding hydrogen bonds, is displayed in Figure 2.12. These increased hydrogen bond weights were used to implement a pharmacophore-directed docking procedure. In an advanced setting (PLANTS protocol II, see 3.5.2), only the weights w_{hb1} and w_{hb3} were increased to 10, while w_{hb2} was set to its default value of 1.

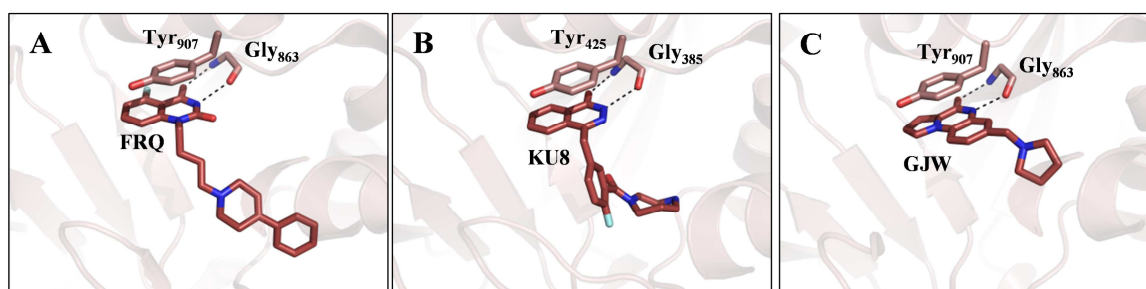
2.6.5 Definition of a correct docking pose

Based on the conformations of co-crystallised inhibitors from *Hs*PARP, *Mm*PARP (Mus musculus, mouse) and *Gg*PARP complexes, a pose was defined as correct if the following features were satisfied:

- Existence of the two essential hydrogen bonds between the docked structure and the conserved glycine residue (e.g. Gly₈₆₃, *Hs*PARP1 numbering)
- Inhibitor core structure being able to exhibit π - π -interactions to the conserved tyrosine residue (e.g. Tyr₉₀₇, *Hs*PARP1 numbering)
- Tail of inhibitor structure does not point towards the protein surface but into the pocket of active site, similar to most *Hs*PARP, *Mm*PARP and *Gg*PARP inhibitors

Definition of a correct docking pose for known and potential new PARP inhibitors

Conformations of co-crystallised *Hs*PARP1 and *Hs*PARP3 inhibitors



docking conformations of co-crystallised *Hs*PARP1 inhibitor 1UK1

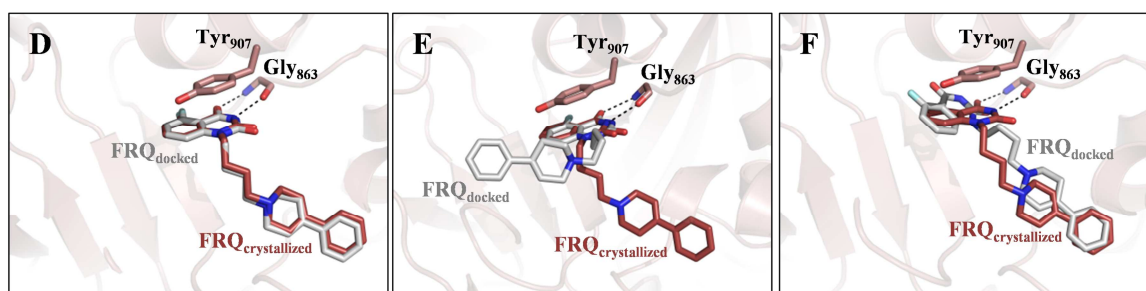


Figure 2.13: Definition of a correct docking pose

A-C: crystallized conformations of PARP inhibitors: A: FRQ in *Hs*PARP1, B: KU8 in *Hs*PARP3, C: GJW in *Hs*PARP1. D: correct docking pose fulfilling all requirements E: incorrect docking pose since tail pointing towards protein surface, F: incorrect docking pose since tail points into active site but shows no hydrogen bonds to Gly₈₆₃

2.7 Methods of probability and inference

2.7.1 Null hypothesis significance testing and statistical power

To compare observed data with a hypothesis whose truth has to be assessed, null hypothesis significance tests (NHST) are performed. The (null) hypothesis $H_0: \mu = \mu_0$, is a statement about a parameter in a population. The results of a test are expressed in terms of a probability. The test measures how well the data and the hypothesis agree. In a NHST, the strength of evidence against the null hypothesis is assessed. To perform NHST, the following steps have to be performed:

Table 2.3: Steps necessary to perform null hypothesis significance testing (NHST)

step	description
1	Choosing a null hypothesis, $H_0: \mu = \mu_0$, where μ is the mean of the population and μ_0 is the mean of the sample data.
2	Choosing a significance level, α , that is commonly chosen to be 0.05, 0.01 or 0.005. If not stated otherwise, α was set to 0.05 when needed.
3	Application of a statistical test to the sample data and calculation of a P -value. The P -value is the probability, that, if H_0 were true, the observed data or more extreme data would be observed.
4	If the P -value is smaller than α , the null hypothesis is rejected and the result is statistical significant at significance level α . Otherwise, H_0 is not rejected and the result not statistically significant at significance level α

If H_0 is rejected, when, in fact, it is true, a Type I error is committed. The probability of rejecting H_0 when it is true, is called the Type I error α . A graphical representation of a Type I error is described in Figure 2.14, A.

An alternative hypothesis $H_1: \mu \neq \mu_0$, can be specified. If H_1 is an exact hypothesis, $H_1: \mu = (\mu_0 + \delta) = \mu_1$, it allows calculation of statistical power. Statistical power is the probability of rejecting H_0 when, in fact, H_1 is true.

If there is a true effect with exact size $(\mu_0 + \delta) = \mu_1$ (which is specified by H_1), statistical power equals the probability, the experiment will find it to be statistically significant. Rejection of $H_1: \mu \neq \mu_0$ when in fact H_1 is true, is called a Type II error. The corresponding probability is called Type II error rate β . Therefore, statistical power is defined as $1-\beta$. This is displayed in , Figure 2.14 B.

Null hypothesis significance testing (NHST)

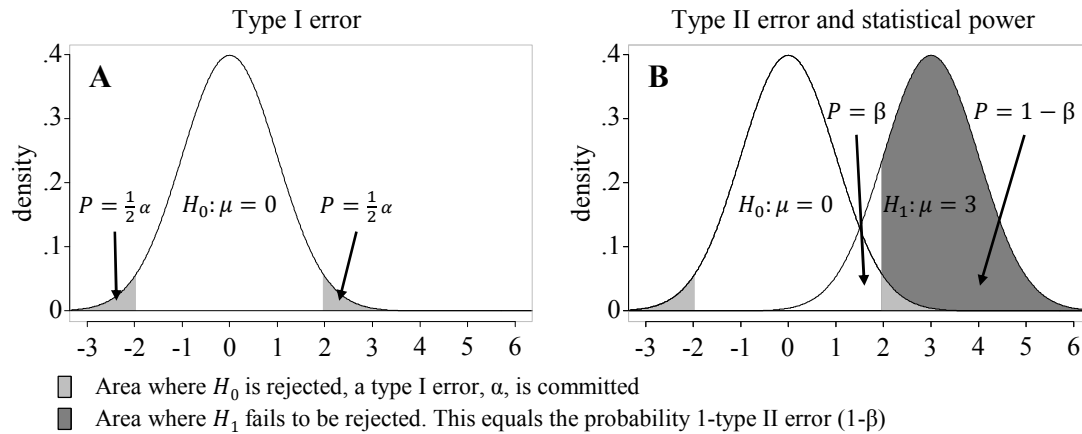


Figure 2.14: NHST and statistical power
 A: type I error, α ; B: type II error, β , and statistical power, $1-\beta$

Since the outcome of a statistical test (rejection or no rejection of H_0 or H_1 , if stated) and therefore the probabilities of committing type I or type II errors, is dependent on the sample mean \bar{x} , sample standard deviation s and the sample size n , one can use power analysis to define thresholds, at which a statistical test has desired type I error rates or statistical power. Also, power analysis can be used to define a minimum sample size, at which a statistical test has sufficient power to identify an underlying effect.

Table 2.4: Relationship between Type I and Type II errors, and statistical power

		Truth about population	
		No treatment effect	treatment effect
conclusion reached in a study	No effect	Correct conclusion	Type II error
	effect	Type I error	Correct conclusion

		Truth about population	
		No treatment effect	treatment effect
conclusion reached in a study	No effect	$P = 1-\alpha$	$P = \beta$
	effect	$P = -\alpha$	$P = 1-\beta$ (stat. power)

2.7.2 One-sample t -test

If a simple random sample (SRS) of size n - having mean \bar{x} and standard deviation s - is drawn from a population having unknown mean μ , the hypothesis $H_0: \mu = \mu_0$ based on an SRS of size n is tested and the one-sample t statistic is computed as follows:

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} \quad (2.6)$$

The standard deviation of the sample mean \bar{x} is defined as the standard error $SE_{\bar{x}} = s/\sqrt{n}$. Let T be a random variable (RV) having a $t(n-1)$ distribution, the P -value for a one-sided test of H_0 against $H_a: \mu > \mu_0$ is $P(T \geq t)$ or a one-sided test of H_0 against $H_a: \mu < \mu_0$ is $P(T \leq t)$. The P -values for a two-sided test of H_0 against $H_a: \mu \neq \mu_0$ is $2 \times P(T \geq t)$.

2.7.3 Unpaired two-sample t -test

If an SRS of size n_1 - having mean \bar{x}_1 and standard deviation s_1 - is drawn from a population having unknown mean μ_1 , and an independent SRS of size n_2 - having mean \bar{x}_2 and standard deviation s_2 - is drawn from a population having unknown mean μ_2 , the hypothesis $H_0: \mu_1 = \mu_2$ based on these two SRS is tested and the two-sample t statistic is computed as follows:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \quad (2.7)$$

The degrees of freedom k , which are used for calculation of P values or critical values t^* for the $t(k)$ distribution, were approximated by R software.²²⁰

2.7.4 Cumulative distribution function (cdf)

Let X be a random variable. The function F or F_X defined by

$$F_X(x) = P\{X \leq x\} \quad -\infty < x < \infty \quad (2.8)$$

is called the cumulative distribution function (cdf) or distribution function of X . $F_X(x)$ expresses the probability that the random variable is less than or equal to x . It is a nondecreasing function and is right continuous.

If X is a discrete random variable, then $F_X(x)$ can be expressed in terms of its probability mass function $p(a)$ by

$$F(a) = \sum_{\text{all } x \leq a} p(x) \quad (2.9)$$

Since X can take on at most a countable number of values, $F(a)$ is the sum of the probabilities $p(a)$ of all values a that are smaller than or equal to x .

If X is a continuous random variable, then $F_X(x)$ can be expressed of its probability density function $f_X(x)$ by

$$F(a) = P\{X \in (-\infty, a]\} = \int_{-\infty}^a f(x) dx \quad (2.10)$$

2.7.5 Pearson's Chi-squared test

Pearson's χ^2 test tests the null hypothesis H_0 that the number of observed events in a simple random sample (SRS) equals the number of events one would observe if the sample was drawn from a specific distribution. The value of the test statistic is equal to:

$$\chi^2 = \sum_{i=1}^n \frac{(o_i - e_i)^2}{e_i} \quad (2.11)$$

Here, χ^2 represents Pearson's cumulative test statistic. This statistic approaches a χ^2 -distribution with n degrees of freedom as the sample size increases. o_i and e_i equal the number of an observed and expected or theoretical frequency and n represents the number of cells in a frequency table. As a result, the value of the test statistic is compared to a χ^2 -distribution with n degrees of freedom. From the value of the test-statistic, a P-value is calculated. If the calculated P-value is smaller than a critical P-value, (e.g. 0.05) H_0 is rejected and one concludes that there is no association between the sample and the theoretical distribution. If the calculated P-value is larger than the critical P-value, than H_0 cannot be rejected and the test-statistic offers no evidence against H_0 .

2.7.6 Binary quantitative structure-activity relationship

2.7.6.1 Binary quantitative structure-activity relationship terminology

Binary quantitative structure-activity relationship (QSAR) correlates the structure of compounds with a binary expression of activity by using molecular descriptors. With the information of activity and molecular descriptors for active and inactive compound structures in a training set, probability distributions for active and inactive compounds are calculated. From those probability distributions, activities of compound structures in a test set can be predicted in a semiquantitative manner. This approach was developed by Labute and colleagues in 1999²²¹ and successfully applied, e.g. on estrogen receptor ligands²²² and tiagabine analogues.²²³ A binary QSAR workflow is implemented in the software package MOE.

Suppose one is given m molecules where each molecule i is described as a vector \mathbf{x}_i being of length n with $\mathbf{x}_i = (x_{i1}, \dots, x_{ij}, \dots, x_{in})$ and $x_{ij} \in \mathbb{R}$. The x_{ij} are called the descriptors of molecule i . Let y_i be an outcome of an experiment for molecule i , e.g. a biological activity, expressed as an IC_{50} or K_i value. The outcome y_i is binary, e.g. $y_i \in \{0, 1\}$. Y denotes a RV with $Y \in \{0, 1\}$ and X denotes a RV over vectors of length n .

In binary QSAR, the conditional distribution $\Pr(Y|X)$ is used to determine the probability, $p(\mathbf{x})$, that a new molecule \mathbf{x}_{new} is active with $\Pr(Y = 1|X = \mathbf{x}_{new})$. Let \mathbf{a} be the prior probability $\Pr(Y = 1)$ and $f(\mathbf{x}, \mathbf{y}) = \Pr(X = \mathbf{x}_{new}|Y = \mathbf{y})$, then by using Bayes' Theorem, $p(\mathbf{x})$ can be written as:

$$p(\mathbf{x}) = \Pr(Y = 1|X = \mathbf{x}) = \frac{f(\mathbf{x}, 1)\mathbf{a}}{f(\mathbf{x}, 1)\mathbf{a} + f(\mathbf{x}, 0)(1 - \mathbf{a})} \quad (2.12)$$

Furthermore, all descriptors X_i are assumed to be mutually independent and having mean 0 and variance 1. After rearranging, the distributions $f_j(\mathbf{x}, \mathbf{y}) = \Pr(X_j = x|Y = \mathbf{y})$ and the prior probability \mathbf{a} have to be estimated. The probability is estimated by the biased Bayes estimate under a uniform prior $\mathbf{a} = (\mathbf{S} + 1)/(\mathbf{m} + 2)$. Here, \mathbf{S} equals the number of actives and \mathbf{m} represents the total number of structures in the data set, where the total number of structures \mathbf{m} consists of the sum of the number of active structures \mathbf{m}_1 and the number of inactive structures \mathbf{m}_0 . This procedure results in formula 2.13.

$$p(x) \approx \left(1 + \frac{m_0 + 1}{m_1 + 1} \prod_{j=1}^n \frac{f_j(x_j, \mathbf{0})}{f_j(x_j, \mathbf{1})} \right)^{-1} \quad (2.13)$$

Furthermore, let $\mathbf{z}_1, \dots, \mathbf{z}_i, \dots, \mathbf{z}_m$ be m samples of a continuous random variable \mathbf{Z} . f can be estimated by accumulating a histogram of observed sample values on a set of \mathbf{B} bins, that are defined by $\mathbf{B} + 1$ numbers, $b_j < b_{j+1}$. Counting the number of observations B_j among m samples into bin $j > \mathbf{0}$ is done by using a δ -function whose density can be replaced by a normal RV with mean \mathbf{z}_i and variance s^2 , giving:

$$B_j = \sum_{i=1}^m \delta(\mathbf{z}_i \in (b_{j-1}, b_j]) = \sum_{i=1}^m \int_{b_{j-1}}^{b_j} \frac{1}{s\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mathbf{z}_i}{s}\right)^2} dx \quad (2.14)$$

By final translation of the normal cumulative distribution function into the error function **erf**, each of the descriptor distributions $f_j(x, \mathbf{y}) = \Pr(X_j = x | Y = \mathbf{y})$ for $\mathbf{y} \in \{\mathbf{0}, \mathbf{1}\}$ and for n descriptors can be modelled. In binary QSAR, two distributions for each descriptor are estimated, one being for the active molecules the other for the inactive molecules.

2.7.6.2 Binary QSAR evaluation

The evaluation of binary QSAR model consists of statements of about the accuracy of the model. The total accuracy denotes the number of observations that were correctly predicted by the model. Two more accuracies denote the number of active and inactive observations that were predicted correctly. The significance of these accuracies is assessed by comparing the observed accuracies with the number of accuracies one would observe if there was no association between the model and the sample.

The null hypothesis H_0 is: There is no association between the model results (of the total accuracy) and the sample. Let $\mathbf{o}_{correct}$ denote the number of agreements and $\mathbf{o}_{incorrect}$ denote the number of disagreements between the model and the sample. Let $\mathbf{e}_{correct}$ and $\mathbf{e}_{incorrect}$ denote the number of agreements and disagreements one would observe if there was no association between the model results and the sample. With increasing sample size, the χ^2 -statistic

$$\chi^2 = \frac{(\mathbf{o}_{correct} - \mathbf{e}_{correct})^2}{\mathbf{e}_{correct}} + \frac{(\mathbf{o}_{incorrect} - \mathbf{e}_{incorrect})^2}{\mathbf{e}_{incorrect}} \quad (2.15)$$

follows a χ^2 -distribution with one degree of freedom. If the associated P-value is smaller than 0.05, then there is strong evidence against H_0 .

To test whether the observed accuracies on actives and inactives are due to chance, another the χ^2 -statistic is calculated. Here, the null hypothesis H_0 is stated: There is no association between the model results (of the active and inactive accuracy) and the sample. Let \mathbf{o}_1 denote the number of agreements in the actives, \mathbf{o}_2 the number of disagreements in the actives, \mathbf{o}_3 the number of agreements in the inactives and \mathbf{o}_4 the number of disagreements in the inactives. Let \mathbf{e}_1 denote the number of agreements in the actives, \mathbf{e}_2 the number of disagreements in the actives, \mathbf{e}_3 the number of agreements in the inactives and \mathbf{e}_4 the number of disagreements in the inactives one would observe if there was no association between the model data and the sample. With increasing sample size, the χ^2 -statistic

$$\chi^2 = \sum_{i=1}^4 \frac{(\mathbf{o}_i - \mathbf{e}_i)^2}{\mathbf{e}_i} \quad (2.16)$$

follows a χ^2 -distribution with three degrees of freedom. If the associated P-value is smaller than 0.05, then there is strong evidence against H_0 . Binary QSAR therefore is evaluated as a special case of the general form described in 2.7.5.

For each principal component, a correlation coefficient is reported, stating how well the active and inactive distributions are correlated. A value of the correlation coefficient of 0 means perfect correlation whereas a value of 1 means that both distributions are perfectly uncorrelated. In addition to that, an RMSE is reported, which is referring to the expected root mean squared error between the active and inactive distributions. Finally, the importance of each descriptor is given, representing the degree to which the descriptor is useful in distinguishing actives from inactives.

2.7.7 Receiver Operator Characteristics

Receiver operator characteristics (ROC) can be used to assess the ability of a procedure to discriminate two classes. An ROC plot is generated by plotting the fraction of true positives tp against the fraction of false positives fp for any given threshold. This results in a curve starting from the point $(0,0)$ and ending up in point $(1,1)$. Any point on the plot corresponds to a certain threshold and according true positive und false positive rate. For a procedure which is highly discriminative, either a strongly increasing curve passing the point $(0,1)$ in ROC space very closely or a very slowly increasing curve passing the point $(1,0)$ in ROC space very closely, is desired. A curve having a similar shape as the direct line from $(0,0)$ to $(1,1)$ does not discriminate the two classes.²²⁴ An example of an imaginary data set representing a chemical characteristic of 5 active and inactive compounds is shown in Figure 2.15

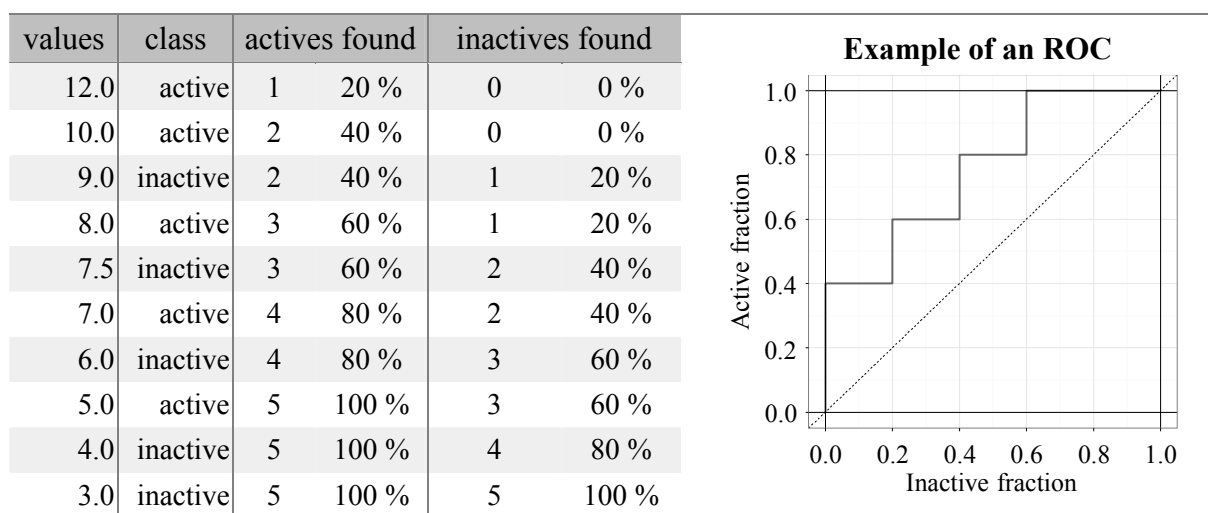


Figure 2.15: Example of an ROC

Left: example data set consisting of 10 imaginary values, sorted in descending order. Right: ROC curve (solid line from $(0,0)$ to $(1,1)$) resulting from example data set; dashed line represents $AUC = 0.5$, indicating no discriminating characteristics

The area under the curve (AUC) represents the probability, that, in the long run, if a member of the active and inactive class are randomly chosen, the procedure will correctly rank the active member above the inactive one. If a member from one class is chosen at random and a member of the other class is chosen at random for many times, the AUC represents the probability that the member of the active class will be selected. Hence, ROC are equivalent in results to Wilcoxon signed-rank test.²²⁵ Furthermore, since ROC are based on fractions rather than absolute numbers, ROC can be compared to each other. That is an advantage over enrichment factors (EF) that depend not only on the numbers of members of the classes, but also on the ratio between both numbers.

2.7.8 Analysis of MD simulations

MD simulations were analysed using the root mean square deviation (RMSD)^{226,227} and the room mean square fluctuation (RMSF). Let \mathbf{a} and \mathbf{b} being atoms in a \mathbb{R}^3 , the position of \mathbf{a} and \mathbf{b} can be described in terms of its x , y and z coordinates, e.g. $\mathbf{a} = (a_x, a_y, a_z)$. With this, the RMSD and RMSF are calculated as follows:

$$\begin{aligned} RMSD(a, b) &= \sqrt{\frac{1}{n} \sum_{i=1}^n ((a_{ix} - b_{ix})^2 + (a_{iy} - b_{iy})^2 + (a_{iz} - b_{iz})^2)} \\ &= \sqrt{\frac{1}{n} \sum_{i=1}^n \|\mathbf{a}_i - \mathbf{b}_i\|^2} \end{aligned} \quad (2.17)$$

$$RMSF(a) = \sqrt{\frac{1}{T} \sum_{t_j=1}^T (x_a(t_j) - \bar{x}_a)^2} \quad (2.18)$$

$$B\text{-factor}(a) = \frac{80}{3} (\pi * RMSF(a))^2 \quad (2.19)$$

RMSD and RMSF are measures of deviations of a set of atoms (e.g. C_α atoms of proteins) from their reference position, which in this case is the mean average position of the C_α atoms. While the RMSD equals the mean average over n atoms for specific time values in relation to a reference time value, the RMSF measures the time-dependent mean average of specific atoms x_i .

Results of MD simulations can be compared to results of X-ray structures because of the relationship between the RMSF of an atom in the MD simulation and the B-factor of the corresponding atom of a protein crystal during X-ray crystallography.

All MD simulations were analysed using either predefined or modified YASARA macros or by R software²²⁰ and the package bio3D (version 2.1)²²⁸

3 Results

3.1 Sequence analysis

Sequence analyses have been performed for the catalytic domains of *Hs*PARP1-3, *At*PARP1-3 and *Gg*PARP1. Table 3.1 summarises general sequence information of the catalytic domains of investigated PARP. Catalytic domain information was extracted from Pfam 27.0 database.²²⁹

Table 3.1: General information about catalytic domains (CD) in selected PARP

	Sequence length	CD start	CD end	CD length
<i>At</i> PARP1	637	286	633	348
<i>At</i> PARP2	983	633	979	347
<i>At</i> PARP3	814	449	801	353
<i>Hs</i> PARP1	1014	662	1007	346
<i>Hs</i> PARP2	583	231	577	347
<i>Hs</i> PARP3	533	182	533	352
<i>Gg</i> PARP1	1011	659	1004	346

In part, the results were compared to conserved residues in Diphtheria Toxin. The catalytic domain-comprising amino acids were taken from Pfam database 27.0.²²⁹ The results are displayed in Figure 3.1. Pairwise sequence analyses (2.2.1) of the catalytic domains of human and *Arabidopsis thaliana* PARPs 1-3, as well as chicken PARP1 (*Gg*PARP1) revealed high sequence identities as well as high sequence similarities between *Hs*PARP and *At*PARP. Upper triangle values in Table 3.2 represent sequence similarities, whereas entries in the lower triangle show sequence identities between two proteins. The complete pairwise sequence alignments are displayed in 5.1.

Table 3.2: Sequence similarities and sequence identities for selected PARP

	<i>At</i> PARP1	<i>At</i> PARP2	<i>At</i> PARP3	<i>Hs</i> PARP1	<i>Hs</i> PARP2	<i>Hs</i> PARP3	<i>Gg</i> PARP1
<i>At</i> PARP1	-	60.2	47.5	65.1	67.4	51.4	65.6
<i>At</i> PARP2	40.4	-	54.1	68.7	64.2	49.5	69.0
<i>At</i> PARP3	28.3	33.1	-	49.9	49.3	40.6	50.6
<i>Hs</i> PARP1	45.7	49.6	28.6	-	68.6	55.0	95.7
<i>Hs</i> PARP2	49.4	43.8	28.6	45.4	-	50.8	67.4
<i>Hs</i> PARP3	35.7	33.8	23.3	35.6	34.0	-	52.9
<i>Gg</i> PARP1	45.7	48.5	29.2	87.0	46.3	35.3	-

upper triangle values represent sequence similarities; sequence identity values shown in lower triangle

The most homologue *At*PARP to *Hs*PARP1 are *At*PARP1 and *At*PARP2, sharing more than 45% sequence identity and more than 65% sequence similarity (bold numbers in Table 3.2). PARP members *At*PARP3 and *Hs*PARP3 show less sequence identity and similarity to *Hs*PARP1-2, *At*PARP1-2 and *Gg*PARP1, respectively.

Multiple sequence analyses (2.2.2) of the catalytic domains of *Hs*PARP1-3, *At*PARP1-3 and *Gg*PARP1 show a conservation of residues that are described as essential for NAD⁺ binding in *Hs*PARP1. Especially the catalytic triad HYE (His₈₆₂-Tyr₉₀₇-Glu₉₈₈, *Hs*PARP1 numbering) is conserved in *Hs*PARP1-3 and *At*PARP1 and *At*PARP2. This is displayed in Figure 3.1. In *At*PARP3 the catalytic histidine is replaced by a cysteine (Cys₆₅₃).³ Interestingly, besides the high conservation of NAD⁺-recognizing residues, there are three residues that are replaced within *At*PARP1-3 and *Hs*PARP1-3. In *Hs*PARP1 one of those is Glu₇₆₃ which is replaced by a positively charged Lys₇₃₅ in *At*PARP2. Also, *Hs*PARP1 Asp₇₆₆ is replaced by glutamate residues in *At*PARP1 (Glu₅₅₅) and *At*PARP2 (Glu₇₃₈). A third exchange within this region might be of importance where the human PARP1 Asp₇₇₀ is replaced by a glutamate in *At*PARP1 (Glu₅₅₉). Both *Hs*PARP1 Asp₇₆₆ and Asp₇₇₀ are implicated in pyrophosphate recognition of NAD⁺.⁷²

Multiple sequence alignment of *At*PARP1-3, *Hs*PARP1-3 and *Gg*PARP1 in relation to residues conserved in Diphtheria Toxin

DiphTox	19	30	34	52	145
	SYHGTK	SI	GIQKP	GFYSTDNKYDAAGYSV	SSVEYI
<i>At</i> PARP3	551 ASAFETVRDIN	651 LWCGSR	662 RHIYKGFIPA	685 AIVCSDAAAEAARYGF	779 EYNEYA
<i>Hs</i> PARP3	283 KDMLLVLDIE	382 LWHGTN	393 AILTSGLRIM	412 GIYFASENSKSAGYVI	511 SQSEYL
<i>At</i> PARP2	734 VKMLEALQDIE	831 LWHGSR	842 GILNQGLRIA	865 GIYFADLVSKSAQYCY	957 MYNEYI
<i>Hs</i> PARP1	762 VEMLDNLLDIE	860 LWHGSR	871 GILSQGLRIA	894 GIYFADMVSKSANYCH	985 LYNEYI
<i>Gg</i> PARP1	762 VQMLDNLLDIE	860 LWHGSR	871 GILSQGLRIA	894 GIYFADMVSKSANYCH	985 LYNEYI
<i>At</i> PARP1	386 IEMVEALGEIE	484 LWHGSR	495 GILSQGLRIA	518 GVYFADMFSKSANYCY	611 LYNEYI
<i>Hs</i> PARP2	331 IQLEALGDIE	426 LWHGSR	437 GILSHGLRIA	460 GIYFADMSSKSANYCF	555 NYNEYI
	. . : : * :	** * : .	: * :	. : . : : : * *	. **

Figure 3.1: Multiple sequence alignment between selected PARP

Amino acids presented in single-letter-code; red: hydrophobic; green: hydrophilic; magenta: positively charged; blue: negatively charged; *: conserved amino acid; :: partially conserved amino acid

3.2 Homology Modelling

Homology modelling was performed in YASARA (2.3.3.1). The crystal structure of the catalytic domain of *Hs*PARP1 (PDB entry 1UK1, including residues Lys₆₆₂ - Thr₁₀₁₁), together with the inhibitor 24, FRQ (Figure 2.10) was used as a single template for modelling the catalytic domains of *At*PARP1 (residues Gln₂₈₆ - His₆₃₇) containing 352 amino acids and *At*PARP2 (residues Ser₆₃₃ - Arg₉₈₃) containing 351 amino acids. Figure 3.2 shows the similarities in the three-dimensional shape between the template structure and the target structures. The catalytic triad HYE which is present in both *At*PARP1 and *At*PARP2 forms the same interactions between the protein and the ligand as it does in *Hs*PARP1. The *Hs*PARP1 Asp₇₆₆-replaced residues Glu₅₅₅ in *At*PARP1 and Glu₇₃₈ in *At*PARP2 are able to exhibit the same interactions with the template inhibitor FRQ (shown in Figure 3.2, E).

A key aspect in obtaining a protein homology model of high quality is the selection of the template or a set of templates. In this work, a single template was used for model generation. In YASARA, a template or multiple templates can be chosen manually or automatically. This is advantageous, if - as it is the case for PARP - multiple crystal (or NMR) structures which would represent suitable templates are available. These templates may differ slightly in three-dimensional space (e.g. in loop regions or through different orientations of amino acids in the active site). Changes in three-dimensional structure also can occur through the presence of small ligands (e.g. inhibitors, co-substrates or co-factors) or molecules that were added to make the crystallisation of the template protein possible. Structural diversities between multiple templates can also occur because of the presence or absence of water molecules in the active site that stabilise the position of the ligand. Selecting any suitable PARP templates from the PDB would probably have resulted in similar models, since there is a high degree of three-dimensional structure similarity in the CD of PARP. More than 30 crystal structures of PARP catalytic domains were deposited at the PDB at the end of 2010. These structures represent CD of *Hs*PARP1-3, *Hs*PARP10, *Hs*PARP14, *Hs*PARP15, *Hs*Tankyrase1-2, *Mm*PARP2, and *Gg*PARP1-2. Sequence identity between those sequences and the catalytic domains of *At*PARP are >30% making all of these structures suitable for homology modeling the CD of *At*PARP. From that it was assumed that, independent from the selected template, structurally highly similar homology models of *At*PARP would be produced.

The selection of a suitable homology modeling template was inhibitor-driven because of the usage for virtual screening for new *At*PARP inhibitors. For the purpose of virtual screening,

the best template would be the one that contributes most information about inhibitor binding. Furthermore, most knowledge about PARP inhibitors exists for human PARP1.

*Hs*PARP1, *At*PARP1 and *At*PARP2 protein models Overviews of tertiary structures and active sites

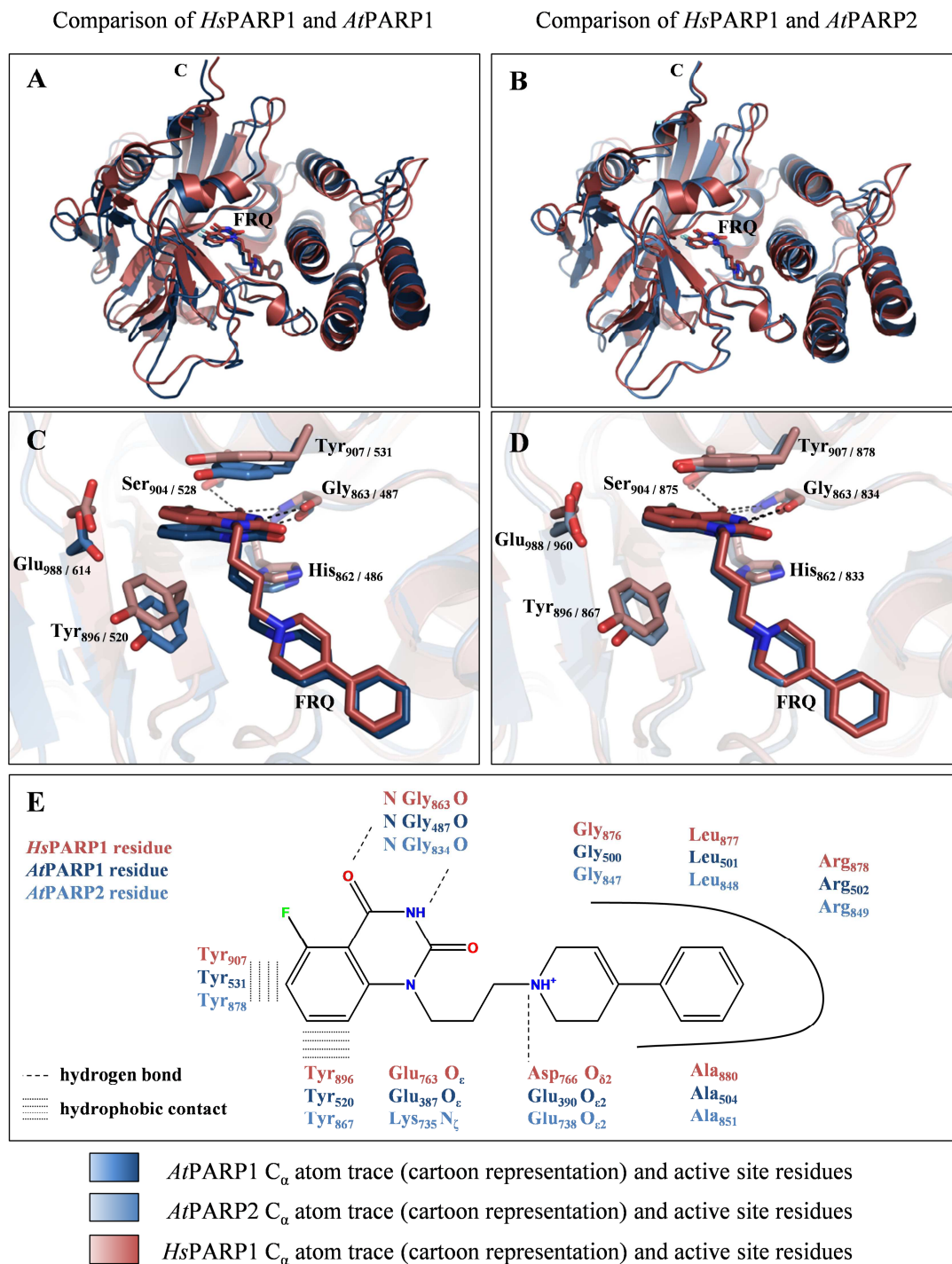


Figure 3.2: Protein models of *Hs*PARP1 and *At*PARP1/2

A and *C*: Comparison of three-dimensional structure and view into active site conserved residues of *Hs*PARP1 and *At*PARP1. *B* and *D*: Comparison of three-dimensional structure and view into active site conserved residues of *Hs*PARP1 and *At*PARP2. *E*: Schematic representation of interactions between modelled ligand FRQ in *At*PARP1/2 and co-crystallised ligand FRQ in *Hs*PARP1 (PDB code 1UK1)

Performing homology modeling with YASARA incorporates the template ligand into the homology model. Therefore, the best choice of template selection was to choose a high-quality *Hs*PARP1 crystal structure containing a large substrate-mimicking structure in its active site.

A major benefit from using a single template in YASARA (e.g. in comparison to using MOE as modelling program) is the transfer of the template ligand into the target structure. During the modelling process in YASARA, amino acid side chain optimisation and unrestrained simulated annealing in an automatically created periodic boundary water box are performed (2.3.3.1) which allow optimal adaptation of the amino acid side chains in active sites with respect to the transferred inhibitor structure. Using this workflow, it could be focussed on the inhibitor during the homology modelling of *At*PARP. Therefore - for the purpose of virtual screening – the crystal structure of *Hs*PARP1 with a co-crystallised quinazolinedione derivative FRQ (24) from PDB entry 1UK1 was used as template because it involves not only a potent (IC_{50} of 60 nM) but also the largest (about 390 g/mol) PARP inhibitor crystallised until 2009. The ligand occupies the donor active site of *Hs*PARP1 ranging from the NA subsite to the hydrophobic pocket that would be occupied by the adenine moiety of NAD^+ . FRQ covers the complete donor site of the active site and bears most information about PARP inhibitors. For these reasons PDB entry 1UK1 was considered an optimal single template for homology modelling.

3.2.1 Evaluation of homology models

Four protein models were subjected to protein evaluation tools. The first model was the crystallographic template structure 1UK1 that was used as a reference for quality assessment. The second model was the energy-minimised model from PDB entry 1UK1. This model was used for docking studies on *Hs*PARP1. Energy minimisation for that model was performed in YASARA using YASARA2 force field. The quality of those two models was compared to the quality of the homology models of *At*PARP1 and *At*PARP2.

The protein evaluation tools check the stereochemical correctness of the models with the Ramachandran plot (RAMPAGE).²³⁰ Errat verifies the overall quality of the models or protein structures based on statistics of non-bonded atom-atom-interactions.²³¹ The quality is expressed as an overall quality factor (with a factor of 100 indicating overwhelming quality). ProSA-web checks for errors in three-dimensional structures and calculates a Z-score as an indicator of native folded proteins. The Z-score is dependent on sequence length and can be

compared to other proteins of similar length.^{232–234} Verify3D investigates the compatibility of three-dimensional structure with its primary sequence.^{235,236} The comparison of the query structure with reference structures allows the calculation of an amino acid-specific score. For assessment of the quality of the homology models in YASARA, a Z-score was provided (2.3.3.1). The results of protein evaluations are displayed in Table 3.3.

Table 3.3: Evaluation of homology models

tool	<i>Hs</i> PARP1 (PDB 1UK1) ¹	<i>Hs</i> PARP1 (optimised) ²	<i>At</i> PARP1 (YASARA) ³	<i>At</i> PARP2 (YASARA) ³
YASARA ^a			-0.791	-1.181
ProSA-web ^b	-9.88	-9.60	-8.94	-8.75
RAMPAGE ^c	339 (97.1 %)	334 (96.0 %)	330 (94.3 %)	338 (96.8 %)
RAMPAGE ^d	8 (2.3 %)	14 (4.0 %)	20 (5.7 %)	9 (2.6 %)
RAMPAGE ^e	0	0	0	2
Errat ^f	98.834 *	97.076 *	94.960 *	98.251 *
Verify3D ^g	0.08-0.77	0.19-0.75	0.03-0.75	-0.15-0.85

¹ crystal structure from PDB; ² optimised in YASARA using steepest descent energy-minimisation
³ homology model created in YASARA as described in 2.3.3.1, based on PDB entry 1UK1 ^a Z-score;
^b Z-score; ^c amino acids in favourable allowed region; ^d amino acids in allowed region; ^e amino acids
in not-allowed region; ^f overall quality; ^g values between; * passed

All four evaluation tools measure slightly lower quality values for the energy-minimised *Hs*PARP1 model than for the PDB-deposited structure. The differences in quality values between both models are small and the values themselves indicate good quality. This classifies the *Hs*PARP1 model which was used for later docking analysis as a model of good quality and suitable for further investigations.

Both homology models of *At*PARP show bigger differences to their PDB template structure than the energy-minimised *Hs*PARP1. The values themselves are acceptable since they are in the range of allowed deviations. The ProSA Z-scores are within the range of Z-scores of protein crystal structures that are of similar length. Also, the ProSA-plots of the PDB structure and the *At*PARP models are of similar shape, especially those with window size 40 (as shown in Figure 3.3). The plots cross the horizontal axes only once and for a small sequence interval. They are below the threshold (being 0) otherwise, which indicates models of good quality. The decreased quality measures can be assigned to gaps in the aligned target and template sequences. Since homology models are built upon the aligned sequences (2.3.3.1), suboptimal scores are mostly found in regions of protein loops or regions of underlying sequence alignment gaps. The YASARA-optimised structures of the *Hs*PARP1 and *At*PARP homology models show similar values in the profile in comparison to the X-ray structure. According to

RAMPAGE, the *At*PARP2 model has two outliers present (Table 3.3 and 5.5.2) which both lie in loop regions of the model, and are not in the proximity of the active site.

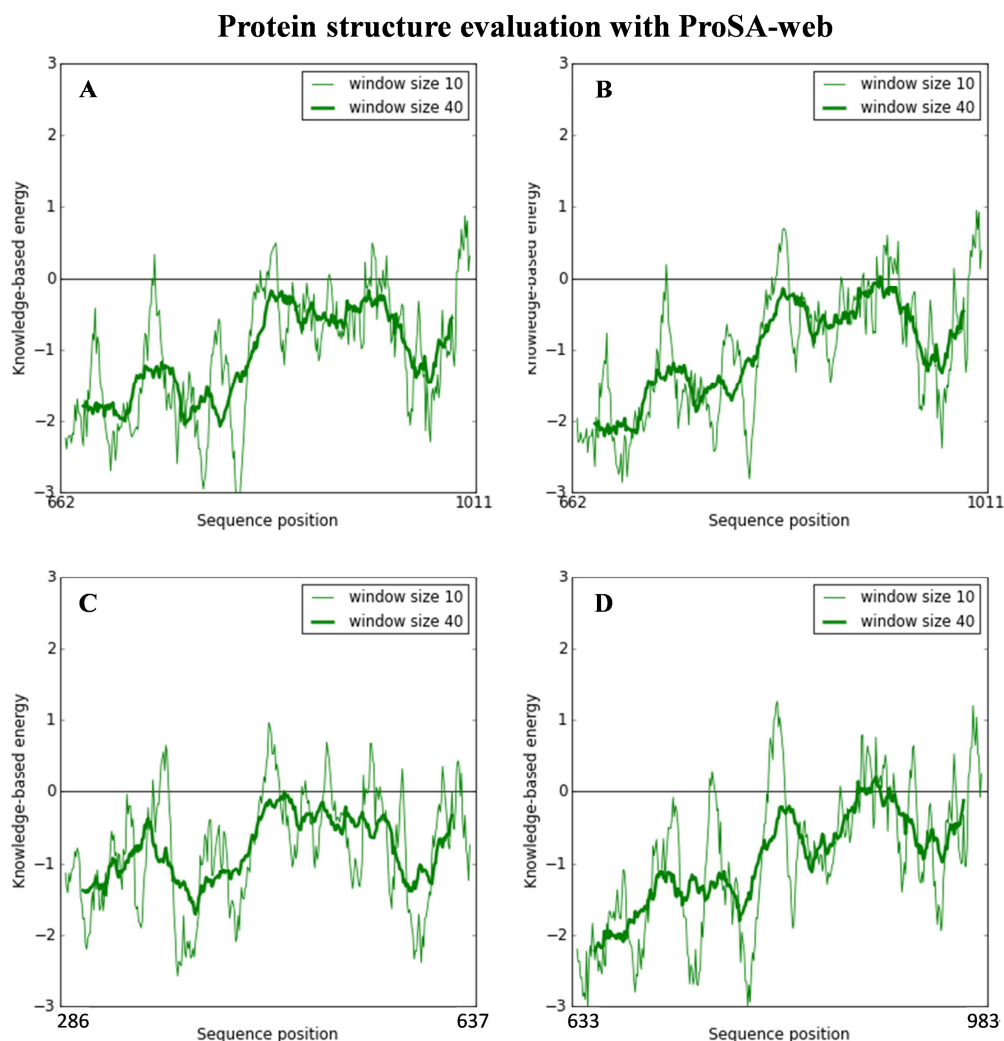


Figure 3.3: Results of ProSA-web

A: the X-ray structure (PDB code 1UK1) of *Hs*PARP1 B: YASARA-Optimised model of *Hs*PARP1 (PDB entry 1UK1) C: homology model of *At*PARP1, D homology model of *At*PARP2

All evaluation tools reveal that the energy-minimised model of *Hs*PARP1 and the homology models of *At*PARP1 and *At*PARP2 qualities that are similar to that of the *Hs*PARP1 X-ray structure. Both ERRAT and Verify3D rate the quality of the *At*PARP models as acceptable (Table 3.3 and 5.4.5).

3.2.2 Model refinement

The result of homology modelling with YASARA is a three-dimensional model of the one-dimensional target sequence. The generated *At*PARP models were partially refined during the modelling via loop optimisation and the simulated annealing of all non-backbone atoms and

finally all atoms including all water molecules in the simulation box (2.3.3.1). In YASARA included is a tool performing a so-called “MD refinement” to improve the quality of the model. In this MD refinement, a 500 ps MD simulation is performed. After every 25 ps, the simulation snapshot is energy-minimised which results in 20 conformations of the protein model with different qualities. To improve the quality of the homology models of *At*PARP1 and *At*PARP2, an MD refinement for each model was conducted. To check if the 20 MD-refined models have improved quality in comparison to the initial model, the protein evaluation tools as described in 3.2.1 were used. Depending on the kind of output of each evaluation tool, the hypothesis was tested, if there was a significant difference (testing with methods described in 2.7.2 and 2.7.3) between the initial homology model and the sample of 20 MD-refined models.

3.2.2.1 RAMPAGE

For *At*PARP1, RAMPAGE found that 18 of 20 MD-refined models had less than 20 amino acids being in the allowed region or outliers as were observed in the initial *At*PARP1 homology model. A similar effect was observed for *At*PARP2, where 19 of 20 MD-refined models had improved quality according to RAMPAGE. The number of amino acids for both *At*PARP1 and *At*PARP2 was significantly reduced during MD-refinement. In contrast to this observation, the number of amino acids during the complete set of MD-refined structures increased significantly. The total number of amino acids during MD-refinement, that were not in the favourable region, increased from 20 to 35 (5.5.2). A similar picture occurred for *At*PARP2, where the number of amino acids in the non-favourable region increased from 9 in the initial model to 29 different amino acids over the period of the 20 snapshots of MD refinement (5.5.2). In each snapshot, a different set of amino acids were detected as being in the non-favourable region which made it difficult to select one of the 20 snapshots as “the best” refined one based and the RAMPAGE results.

3.2.2.2 ProSA-web and Errat

For the tools ProSA and Errat, the hypothesis was tested if there was a significant difference between the values of the initial model in comparison to the values of the 20 MD-refined models. The difference was declared significant if the initial score was not included in the 95% confidence interval of the 20 MD-refined model score which is equivalent to the result of a two-sided unpaired Student’s t-test at significance level $\alpha=0.05$. (2.7.3) For *At*PARP1, the Z-score of ProSA-web offers overwhelming evidence that the MD refinement does not improve the model quality whereas the overall quality score of Errat shows strong evidence

that the quality of the model has improved during MD-refinement (5.5.1). For *AtPARP2*, there is overwhelming evidence that the Z-score decreased during the MD refinement which corresponds to an improvement of *AtPARP2*'s model quality during MD refinement. The overall quality scores of Errat during MD-refinement show no evidence of quality improvement (5.5.1).

3.2.2.3 Verify3D

For Verify3D and each amino acid of *AtPARP1/2*'s primary sequence, the minimum and maximum of the average 3D-1D score for all 20 MD-refinement snapshots were used to define a corridor to which the corresponding average 3D-1D score of the initial homology model was compared. For *AtPARP1*, in the initial model, more than 21% of the catalytic domain show equal or better scores than the best scores of the MD-refinement snapshots whereas only around 5% show equal or worse scores than the worst scores of the MD-refinement snapshots (indicated by green or red bars in figures of section 5.5.3, respectively). The remaining 73% of the initial model scores lie within the range of values for the MD-refinement snapshots (5.5.3). Similar results are obtained for *AtPARP2* where more than 11% of the catalytic domain show equal or better scores than the best scores of the MD-refinement snapshots whereas around 13% show equal or worse scores than the worst scores of the MD-refinement snapshots. The remaining 75% of the initial model scores lie within the range of values for the MD-refinement snapshots (5.5.3). These results strongly indicate that there is no difference between the 3D-1D score profiles of the initial model and the MD-refinement snapshots. A summary of the results of four different evaluation tools to compare the quality of the initial homology models with 20 MD-refined models is shown in Table 3.4.

Table 3.4: Summary of *AtPARP1* and *AtPARP2* model refinement results

	ProSA-web ^a	RAMPPAGE ^b	Errat ^c	Verify3D ^g
<i>AtPARP1</i>	Quality not improved	No difference	Quality improved	No difference
<i>AtPARP2</i>	Quality improved	No difference	No difference	No difference

^a Z-score; ^b amino acids in allowed region and classified as outliers; ^c overall quality score; ^g 3D-1D score profile

These results provide strong evidence that the usage of MD refinement with YASARA does not significantly improve the quality of the homology models of *AtPARP1* and *AtPARP2*. Besides that, the initial models were of quality sufficient for virtual screening. The differences between the evaluation tool outputs of the initial model and the crystal structure template 1UK1 are small enough such that the initial model can be rated as being of equal quality as the crystal structure.

3.3 Investigation of protein stability

To investigate the protein stability of *At*PARP1 models, a homology model of the catalytic domain of *At*PARP1 was built in YASARA based on the template structure 2PAX having the inhibitor 4AN (Figure 2.2, 16) co-crystallised in the active site. For a first model to investigate, the inhibitor 4AN was removed. This model was named unligated *At*PARP1. For a second model, 4AN was kept in the active site. This model was named 4AN-ligated *At*PARP1. Both models were subjected to MD simulations in YASARA as described in 2.3.3.2. After MD simulations, the RMSDs of C α atoms for both models were calculated (2.7.8 and 5.11.2) to find time points where simulations were at equilibrium.

Table 3.5: Summary statistics for MD simulations

T (K)	RMSD for unligated <i>At</i> PARP1				RMSD for 4AN-ligated <i>At</i> PARP1			
	1 st Qu.	mean	median	3 rd Qu.	1 st Qu.	mean	median	3 rd Qu.
297.9999	2.173	2.285	2.266	2.388	2.110	2.277	2.280	2.474
298.0000	2.270	2.463	2.421	2.603	2.117	2.386	2.414	2.727
298.0001	2.021	2.126	2.107	2.225	2.101	2.316	2.287	2.477

T (K): temperature in Kelvin; 1st Qu.: first quartile; value at which 25% of all values are below; 3rd Qu.: third quartile; value at which 75% of all values are below

Descriptive statistics for three independent unligated *At*PARP1 and 4AN-ligated *At*PARP1 MD simulations reveal that all median and mean RMSD values are below 2.5 Å (Table 3.5). These values, together with loess function²³⁷ applied onto RMSD values of each simulation, were used to investigate equilibration of MD simulations. Equilibrations were reached after 5 ns (after 1000 simulation snapshots) for all three unligated *At*PARP1 simulation; equilibrations were reached after 10 ns (at 297.9999 K) and 12.5 ns (at 298.0000 and 298.0001 K) for 4AN-ligated *At*PARP1. (Figure 3.4, represented as grey background). RMSF calculations (2.7.8) were performed for all snapshots after equilibration time.

3.3.1 Overall shape of B-factor and C α distributions

To investigate the fluctuations of C α atoms over those periods and the influence of binding of the inhibitor 4AN in the active site of *At*PARP1 on these fluctuations, RMSF values were calculated for the unligated and 4AN-ligated *At*PARP1, as described in 2.7.8. The hypothesis of similar shapes of B-factor (for *Gg*PARP1) and RMSF value (for *At*PARP1) distributions was tested. *Gg*PARP1 B-factors were extracted from the PDB files 2PAX (4AN-ligated *Gg*PARP1) and 2PAW (unligated *Gg*PARP1).⁷² The distributions of the calculated RMSF

values and experimentally determined B-factors are displayed in Figure 3.5. As expected, higher B-factors occur in regions of loops or less-structured regions of the protein. Major peaks in the B-factor distributions are visible in the region of residues 780, 810 and 830 (GgPARP1 numbering).

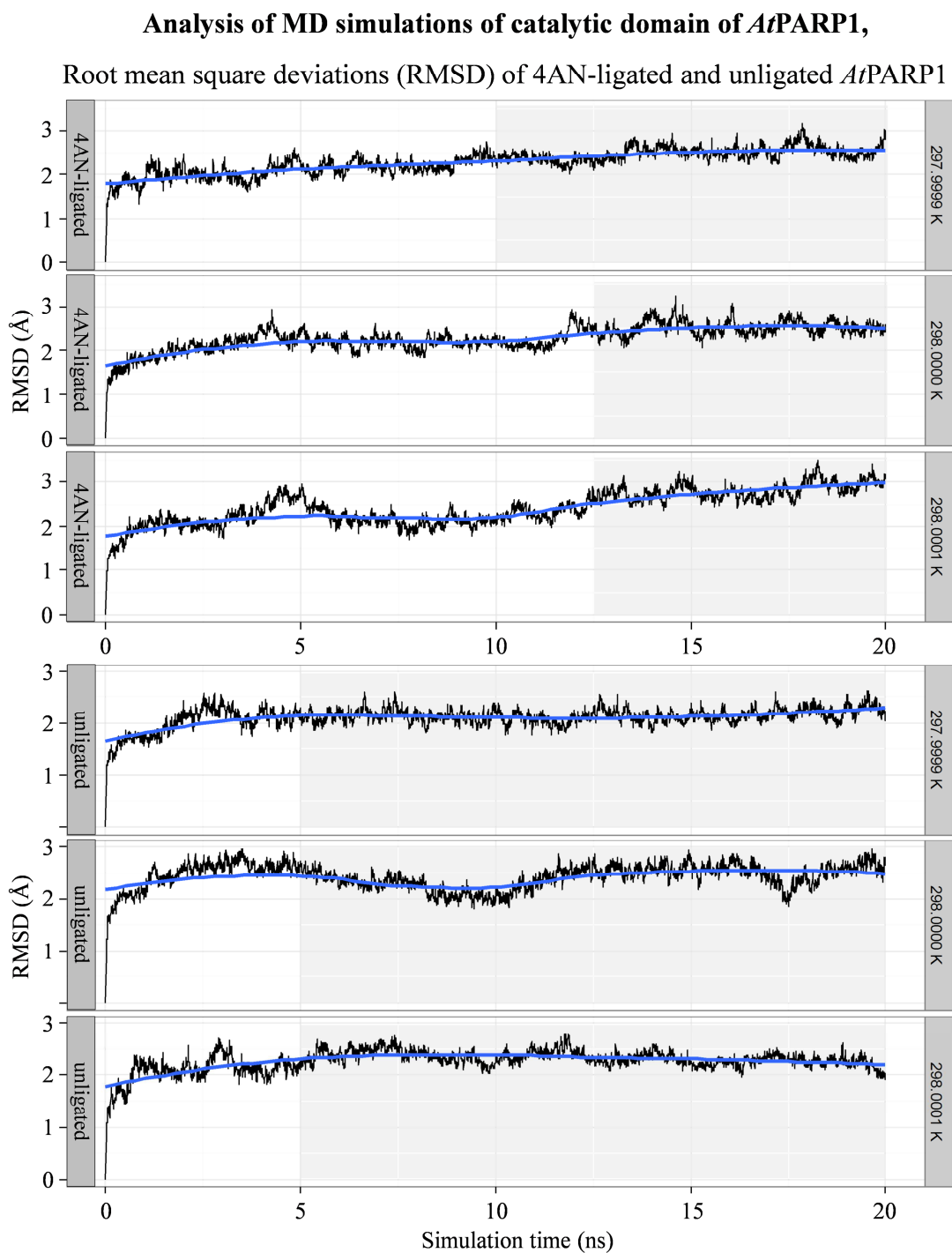


Figure 3.4: Analysis of MD simulations with unligated and 4AN-ligated AtPARP1 I MD simulations as triplicates at three temperatures top 3 graphs::4AN-ligated AtPARP1 simulations at 297.9999 K, :298.0000 K and :298.0001 K. bottom 3 graphs::unligated AtPARP1 simulations at 297.9999 K, :298.0000 K and :298.0001 K. Lowess function (blue line) was applied on RMSD values (black line) to estimate equilibrated MD; equilibrated MD simulation periods are indicated as grey backgrounds.

Furthermore, peaks occur around residues 890, 910, 940 and 980. Corresponding peaks are also visible for the RMSF values of *At*PARP1. There are no regions in *Gg*PARP1 leading to a specific pattern, that is not occurring in the *At*PARP1 model. The overall shape of the distributions is the same and provides little evidence against the hypothesis, therefore the hypothesis is not rejected; the shapes can be regarded as similar. In this specific setting, comparison of MD simulations are sensitive enough to examine C_{α} and B-factor distribution analysis.

B-factor and RMSF distribution of catalytic domains of PARP1,

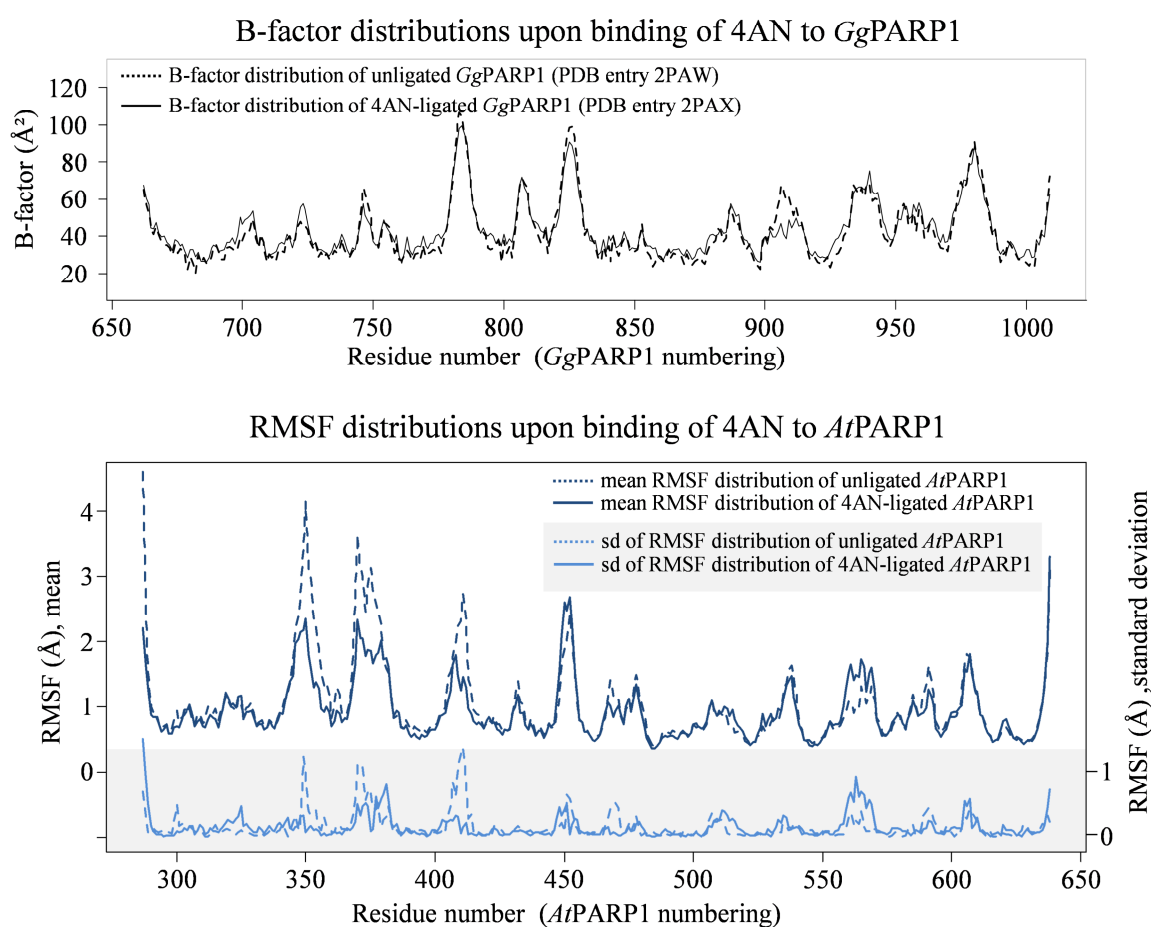


Figure 3.5: Analysis of MD simulations with unligated and 4AN-ligated *At*PARP1 II
Comparison RMSF and B-factor distributions in *Gg*PARP1 and *At*PARP1 top: B-factor distributions in *Gg*PARP1 upon binding of 4AN (modified from Ruf⁷²). bottom: RMSF distributions in *At*PARP1 upon binding of 4AN. upper part (white background) represents mean RMSF data; lower part (grey background) shows standard deviation of RMSD data

3.3.2 Local structural shifts upon inhibitor binding

The second hypothesis - regarding the stabilising effect upon inhibitor binding - was tested by comparing the RMSF values of the two sets of MD simulation triplicates around *At*PARP1 residues Ser₅₂₈ and Asn₅₃₅ (comparison of dotted and solid lines in Figure 3.6). It was investigated whether a decrease in RMSF values is observed, similar to the observation of a decrease in B-factors from 46-57 to 31-36 Å² - corresponding to a 25-43% decrease - upon inhibitor binding in *Gg*PARP1.

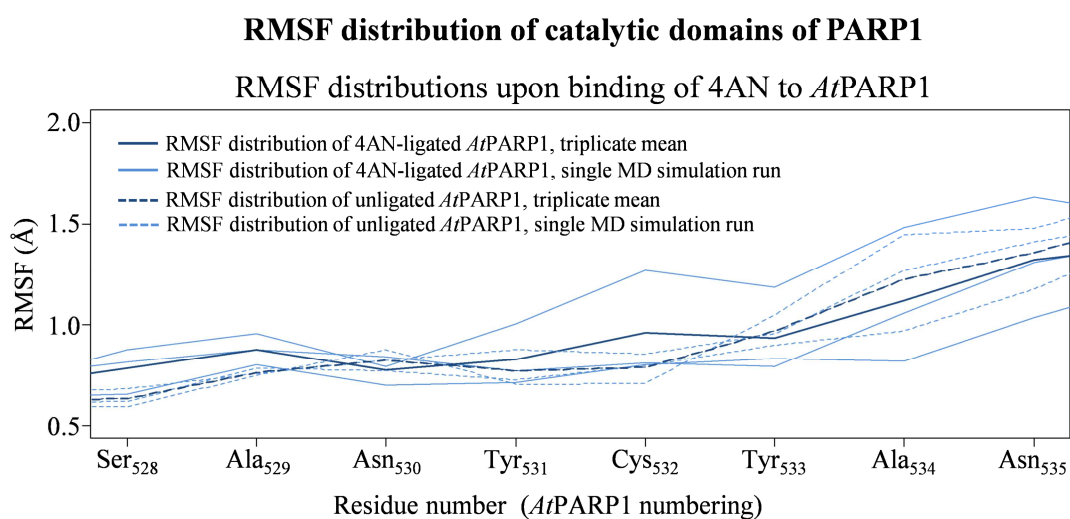


Figure 3.6: Analysis of MD simulations with unligated and 4AN-ligated AtPARP1 III
Comparison of RMSF distributions and AtPARP1; solid dark blue line and dashed dark blue lines represent mean RMSD values in 4AN-ligated and unligated AtPARP1, respectively; solid light blue line and dashed light blue lines represent mean RMSD values in 4AN-ligated and unligated AtPARP1, respectively

This stabilising effect upon inhibitor binding was not observed for *At*PARP1. In fact, in conducted MD simulations, the RMSF values are higher in the corresponding region for the 4AN-ligated protein. With regard to the RMSF of Tyr₅₃₁, the mean RMSF value increases from 0.768 to 0.827, corresponding to an increase of 7.6% upon binding of 4AN in relation to unligated *At*PARP1 (Figure 3.6, dotted and solid dark blue lines). There is serious evidence against this hypothesis which therefore has to be rejected. Based upon the MD simulations, conducted under conditions described in 2.3.3.2, the stabilising effect upon inhibitor binding of the loop around *At*PARP1 Tyr₅₃₁ cannot be proven. From the experiments one cannot assess whether this stabilising effect does not exist at all in *At*PARP1 or whether an existing local effect is just not captured using this experimental setting.

3.4 Investigation of positions of natural substrates

The investigation of the binding mode of natural substrates is essential since it is the basis for subsequent virtual screening. If the predicted (or already elucidated) binding mode of a protein's natural substrate cannot be shown for the target under investigation, there is strong evidence that the modelled active site is in need of improvement. Only if there is enough evidence that the binding mode of the protein's natural substrate can be captured and can be well explained by the model, than the model is of sufficient quality. Regarding *At*PARP1, it is of importance that the binding mode of NAD^+ is modelled correctly because in subsequent virtual screening for competitive inhibitors one can assume that one is searching for structures with similar three-dimensional shape as naturally bound NAD^+ .

There is no crystallographic model available containing PARP's substrate NAD^+ in the donor site of the catalytic domain of PARP. A single X-ray structure was solved having an NAD-analogue bound in the acceptor site of *Gg*PARP1. ADP-ribosylating bacterial toxins use NAD^+ as substrate as it is supposed that PARP bind NAD^+ in a similar manner. To investigate whether NAD^+ is bound in *At*PARP1 in a similar way as it is proposed for *Gg*PARP1 and bacterial toxins, and also whether its conformation is in a favourable position to form a glycosilic bond between the donor and acceptor riboses, direct approaches of homology modelling were not applicable. As a consequence, an indirect approach based on templates of *Gg*PARP1 and ADP-ribosylating toxins has been used to model the natural substrate binding in *At*PARP1. (2.4.2)

3.4.1 Homology model of natural substrate-bound *At*PARP1

The ProBiS algorithm found 39 structures having an active site being structurally similar to the active site of NAD^+ -bound Diphtheria Toxin (PDB entry 1TOX). The six best structures according to active site similarity to Diphtheria Toxins (measured in ProBiS as Z-scores) are the crystal structures of Diphtheria Toxin itself (hits 1 and 2, PDB entries 1DDT and 1DTP, Z-scores of 4.77 and 4.43), Exotoxin A (hits 3 to 5, PDB entries 1XK9, 1XKP and 3Q9O, Z-scores from 3.04 to 2.67) and Choline Toxin (hit 6, PDB entry 3ESS, Z-score of 2.47). These hits prove the ability to find conserved active sites because all these bacterial toxins belong to the enzyme class of ADP-ribosyl transferases.

The highest ranking non-bacterial toxin with ADP-ribosyl transferase activity (7th best hit with a Z-score of 2.44) was the crystal structure of *Gg*PARP1 (PDB entry 1A26) having the

NAD⁺-analogue carba-NAD crystallised in the acceptor site. Six criteria, as part of the ProBiS Algorithm, indicate that the active sites of 1A26 and 1TOX are similar because there is overwhelming evidence against the hypothesis that the calculated local alignment was calculated by chance. All ProBiS criteria with thresholds and observed values for 1TOX and 1A26 are given in Table 3.6.

Table 3.6: ProBiS results

ProBiS result	criteria	observed	ProBiS result	criteria	observed
Z-score	> 2.00	2.44 *	Vertices	> 10	39 ***
Local alignment score	1-10	7.36 **	RMSD	< 2.0 Å	0.7 Å ***
E-value	< 1.0*10 ⁻⁴	3.6*10 ⁻¹² ***	Surface vector angle	< 90°	0.53° ***

* pairwise alignment in top 1% of all alignments in database ** structurally conserved active sites *** overwhelming evidence against the hypothesis that the calculated local alignment match may have occurred by chance

After superpositioning the crystal structures of 1TOX and 1A26 by ProBiS, all non-ligand atoms of 1TOX were deleted, resulting in a model of the crystal structure of the CD of GgPARP1 having NAD⁺ bound in the donor site and CNA bound in the acceptor site. Both superimposed crystal structures and the positions of NAD⁺ and CNA in the superposed active sites are shown in Figure 3.7.

In both active sites, 17 amino acids were calculated to be conserved (listed in 5.8). The results of conserved amino acids between GgPARP1 and Diphtheria Toxin are in agreement with the superposed structures that were inspected visually. Figure 3.7 D-F shows the superposed active sites of both structures together with the location of 9 residues calculated to be conserved.

This structure was used as a single template to homology model the CD of AtPARP1 using YASARA (2.3.3). The homology model of AtPARP1 included both NAD and CNA in the active site. As a next step the model was subjected to 3 independent 20 ns MD simulations in YASARA as described in 2.3.3.2 The simulation equilibrated after 1000 snapshots, that equal 5.0 ns (5.11.3). The 3000 snapshots of the equilibrated MD simulation were used to analyse the interactions between a) AtPARP1 and the substrates and b) the interactions between both substrates themselves. Since the template structures 1TOX and 1A26 contain PARP's substrates (or substrate analogues), the hypotheses were tested if interactions that are observed in crystal structures are found in the MD simulation with AtPARP1, too or in a similar manner.

Superposition of active sites of Diphtheria Toxin and GgPARP1 by ProBiS
Superposition of conserved active site and catalytic residues

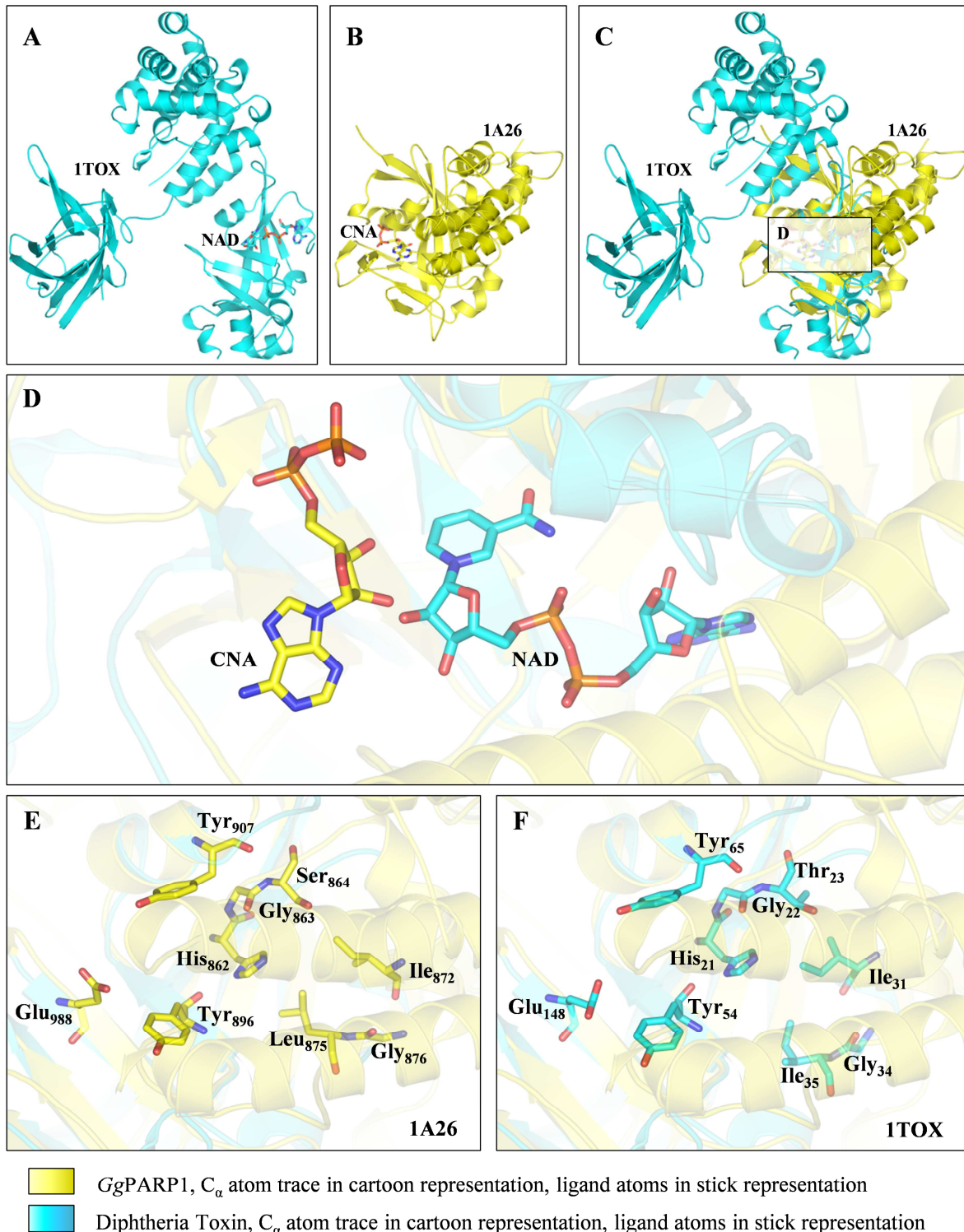


Figure 3.7: ProBiS results I - Diphtheria Toxin and GgPARP1 active sites superposition

A: X-ray structure of Diphtheria Toxin (1TOX), B: X-ray structure of GgPARP1 (1A26), C: superpositioning of both structures using ProBiS algorithm D: View into active site of superposed X-ray structures showing natural substrates of AtPARP1 in close position to each other E: view into active site of GgPARP1 and conserved residues F: view into active site of Diphtheria Toxin and conserved residues

3.4.2 Positioning of the nicotinamide moiety of NAD⁺

During the MD simulations, the nicotinamide moiety of NAD⁺ is held in position tightly by two essential hydrogen bonds. These are formed between the N₇N of NAD⁺ and the backbone oxygen of Gly₄₈₇, as well as the O₇ of the nicotinamide, N₇O, and the backbone nitrogen of Gly₄₈₇, and stabilise the nicotinamide as observed in the majority of the MD simulation times (Table 3.7, interactions NA₁ and NA₂). The hydrogen bonding interaction between the O_γ of Ser₅₂₈ and O₇N of NAD⁺ could not be observed in a single simulation (Table 3.7, interactions NA₆). Despite this interaction being reported to be essential for HsPARP1⁶⁶, crystal structures do exist that contain a nicotinamide-analogue inhibitor in the active site of HsPARP1 and lack this interaction, (e.g. 1UK1¹⁵⁷). Also, in the crystal structure of DT¹²¹, the serine is exchanged by an alanine (Ala₆₂, DT numbering) which does not exhibit this hydrogen bond to the nicotinamide moiety of NAD⁺. This suggests that the hydrogen bond mediated by Serine is of minor importance compared to the hydrogen bonds in that Gly₄₈₇ is involved.

Table 3.7: Results of MD simulations - comparison with experimental data I

interacting atoms in			MD simulation analysis			experimental results		
			distance (Å) or h-bond frequency			GgPARP1	DT	HsPARP1
NAD	AtPARP1		297.9999 K	298.0000 K	298.0001 K	Ref: 72	Ref: 121	Ref: 238
NA _{1_d}	NAD N ₇ N	Gly ₄₇₈ O	2.94 ± 0.16	3.06 ± 0.20	4.36 ± 0.91	2.9	3.0-3.1	2.9
NA _{1_h}	NAD N ₇ N	Gly ₄₇₈ O	19.2 %	82.6 %	97.9 %	yes	yes	yes
NA _{2_d}	NAD N ₇ O	Gly ₄₇₈ N	3.02 ± 0.19	2.94 ± 0.13	3.25 ± 0.34	2.7	2.8-2.9	2.7
NA _{2_h}	NAD N ₇ O	Gly ₄₇₈ N	73.4 %	99.2 %	96.5 %	yes	yes	yes
NA ₃	NAD C ₃ N	Tyr ₅₃₁ C _γ	3.88 ± 0.23	4.27 ± 0.25	4.12 ± 0.28	3.9	3.7-3.8	4.4
NA ₄	NAD C ₆ N	Tyr ₅₃₁ C _ζ	3.50 ± 0.22	3.81 ± 0.24	3.62 ± 0.27	3.8	3.7-3.8	4.4
NA ₅	NAD C ₃ N	Tyr ₅₂₀ C _β	4.96 ± 0.53	3.65 ± 0.18	3.82 ± 0.25	3.9	3.9-4.0	4.1
NA _{6_d}	NAD N ₇ O	Ser ₅₂₈ O _γ	6.23 ± 0.52	4.48 ± 0.53	6.58 ± 1.10	2.7	Ser→Ala	3.2
NA _{6_h}	NAD N ₇ O	Ser ₅₂₈ O _γ	0.0 %	0.0 %	0.0 %	yes	no	yes/ no ¹⁵⁷
NA ₇	Tyr ₅₂₀ O _η	Phe ₅₁₅ N	98.0 %	92.8 %	99.4 %	yes	no	yes
NA ₈	Tyr ₅₂₀ O _η	Gly ₅₁₆ O	95.3 %	53.4 %	93.8 %	yes	yes	yes

Abbreviations: yes: presence of hydrogen bond; no: no hydrogen bond present; Ser→Ala: serine present instead of alanine

The NA moiety of NAD⁺ is furthermore stabilized by the presence of two tyrosine side chains (Tyr₅₂₀ and Tyr₅₃₁). The distances between NA atoms C₃N and C₆N and atoms of the tyrosine side chains either indicate hydrophobic contacts (via π-π interactions) between the side chains and the ring system of NA or steric hinderance between both ring systems (Table 3.7, interactions NA₃ - NA₅). The corresponding observed distances and interactions are in a

similar range to those that have been observed in Diphtheria Toxin, *Hs*PARP1 or NAD^+ modelling studies (Table 3.7).

Analysis of MD simulation

Nicotinamide positioning and π - π stacking

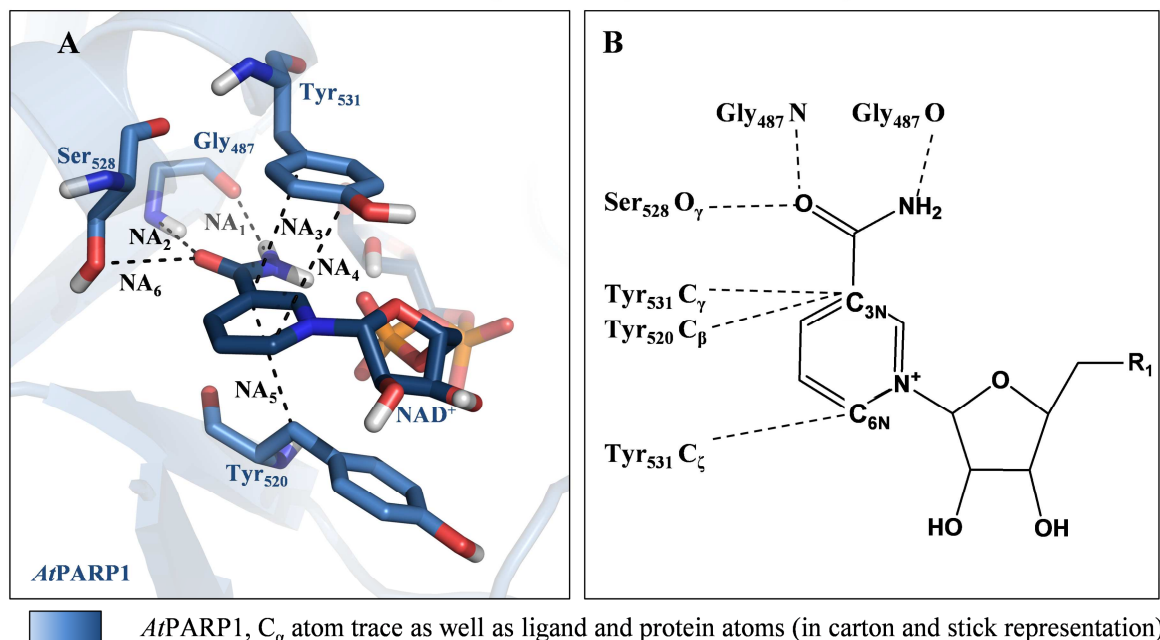


Figure 3.8: Positioning of the nicotinamide moiety of NAD^+ in *At*PARP1

A: view into active site of *At*PARP1; non-polar hydrogens omitted; dotted lines and NA_1 - NA_6 represent interactions or contacts described in Table 3.7. B: schematic representation of NA positioning in *At*PARP1, R_1 = adenosine diphosphate of NAD^+

The side chain of Tyr_{520} itself is fixed via two hydrogen bonds between the O_η to the backbone nitrogen of Phe_{515} and the backbone oxygen of Gly_{516} (interactions NA_7 and NA_8 in Table 3.7, respectively; interactions omitted in Figure 3.8). This fixation is also present in other ADPRT.^{72,121,238} Tyrosine 520 is therefore contributing to the stacking of the NA moiety of NAD^+ . This stacking of the NA is essential, as mutations of both tyrosines into asparagines in *Gg*PARP1 result in a reduced enzyme activity of 15% and 1.1%, respectively.³¹ Asparagine, in contrast to tyrosine, is unable to exhibit hydrophobic interactions.

From the interactions that were observed in the MD simulation of *At*PARP1 one can conclude that the NA moiety of NAD^+ is positioned and stabilised in an equivalent manner as it is observed in crystal structures of different ADPRT.

3.4.3 Binding of the adenine moiety of the donor structure NAD⁺

The adenosine moiety of NAD⁺ is predicted to be held in position via two hydrogen bonds (backbone oxygen of Gly₃₄ in DT or Gly₈₇₆ in *Hs*PARP1 and the backbone nitrogen of Gln₃₆ in DT or Arg₈₇₈ in *Hs*PARP1, respectively). Analogue interactions were observed for Arg₅₀₂N (Table 3.8, interaction AD₁) and for Gly₅₀₀O (Table 3.8, interaction AD₂) in the production MD simulation in *At*PARP1. This indicates high flexibility of the adenine moiety since the hydrogen bond frequencies were relatively low and not stable throughout the MD simulations

The NAD⁺ adenine ring is also fixed through non-polar interactions. The side chains of the hydrophobic residues Ile₄₉₆ and Leu₅₀₁ are observed in favorable distances to establish hydrophobic contacts (Table 3.8, interactions AD₃ and AD₄). Analogue observations were also observed in modelling studies of NAD⁺ and the crystal structure of DT. In addition to those interactions, there is evidence that the side chain of Arg₅₀₂ contributes to the fixation of the moiety. By investigating the distance from the guanidinium carbon atom C_{5A} of NAD⁺ to the C_ζ of Arg₅₀₂ (Table 3.8, interaction AD₅), it was observed that the guanidinium group could potentially hinder the adenine ring of NAD⁺ from larger movements.

Table 3.8: Results of MD simulations - comparison of experimental data II

	interacting atoms in		MD simulation analysis			experimental results	
	NAD ⁺	<i>At</i> PARP1	distance (Å) or h-bond frequency			<i>Gg</i> PARP1 ⁷²	Diph. Tox. ¹²¹
			297.9999 K	298.0000 K	298.0001 K		
AD _{1d}	N _{1A}	Arg ₅₀₂ N	3.28 ± 0.27	3.24 ± 0.22	5.12 ± 1.42	2.7	2.9-3.2
AD _{1h}	N _{1A}	Arg ₅₀₂ N	20.0 %	18.6 %	39.0 %	H-bond	H-bond
AD _{2d}	N _{6A}	Gly ₅₀₀ O	4.23 ± 0.45	4.01 ± 0.39	5.04 ± 1.43	3.0	2.7-2.9
AD _{2h}	N _{6A}	Gly ₅₀₀ O	55.9 %	44.3 %	10.2 %	H-bond	H-bond
AD ₃	C _{5A}	Ile ₄₉₆ C _{σ1}	4.85 ± 0.40	4.15 ± 0.28	4.26 ± 0.48	hydrophobic	3.8-4.0
AD ₄	C _{2A}	Leu ₅₀₁ C _{σ2}	3.85 ± 0.27	3.74 ± 0.29	4.96 ± 0.86	hydrophobic	3.9-4.1
AD ₅	C _{5A}	Arg ₅₀₂ C _ζ	4.68 ± 0.36	4.13 ± 0.45	5.27 ± 0.77	---	---
AD ₆	O _{2B}	His ₄₈₆ N _{ε2}	0.1 %	13.2 %	55.1 %	yes	2.8
AD ₇	Ser ₄₈₈ O _γ	His ₄₈₆ N _{ε2}	21.1 %	80.6 %	4.6 %	no	---
AD ₈	O _{2B}	Ser ₄₈₈ O _γ	0.0 %	50.6 %	2.9 %	2.8	2.7-2.9

Abbreviations: H-Bond: presence of hydrogen bond; hydrophobic: hydrophobic contact, no distance given; ---: no equivalent interaction

In *Hs*PARP1 the adenine ribose of NAD⁺ is predicted to interact with the γ oxygen of a serine residue and an analogue interaction is observed between the O_{γ1} of Thr₂₃ in the crystal structure of Diphtheria Toxin with O_{2B} of NAD⁺. In *At*PARP1, the hydrogen bond between the O_γ of Ser₄₈₈ and 2' oxygen of the adenosine ribose of NAD⁺ is observed in only one of

three simulations (Table 3.8, interaction AD₈). But there might be additional interactions between the 2' oxygen of the adenosine ribose and the protein that might stabilise the position of the ribose: First, the side chain of serine 488 binds to the N_{ε2} of His₄₈₆ (Table 3.8, interaction AD₇), while His₄₈₆ N_{ε2} hydrogen bonds to O_{2B} of NAD⁺ (Table 3.8, interaction AD₆). This indicates the presence of a triangular hydrogen bonding network that stabilises the adenosine ribose of NAD⁺. An view into the active site of AtPARP1 with regard to the adenine moiety is shown in Figure 3.9, A, schematic representation of the interactions AD₁ – AD₈ are shown in Figure 3.9, B

Analysis of MD simulations of the catalytic domain of AtPARP1

Binding of adenine moiety of NAD⁺

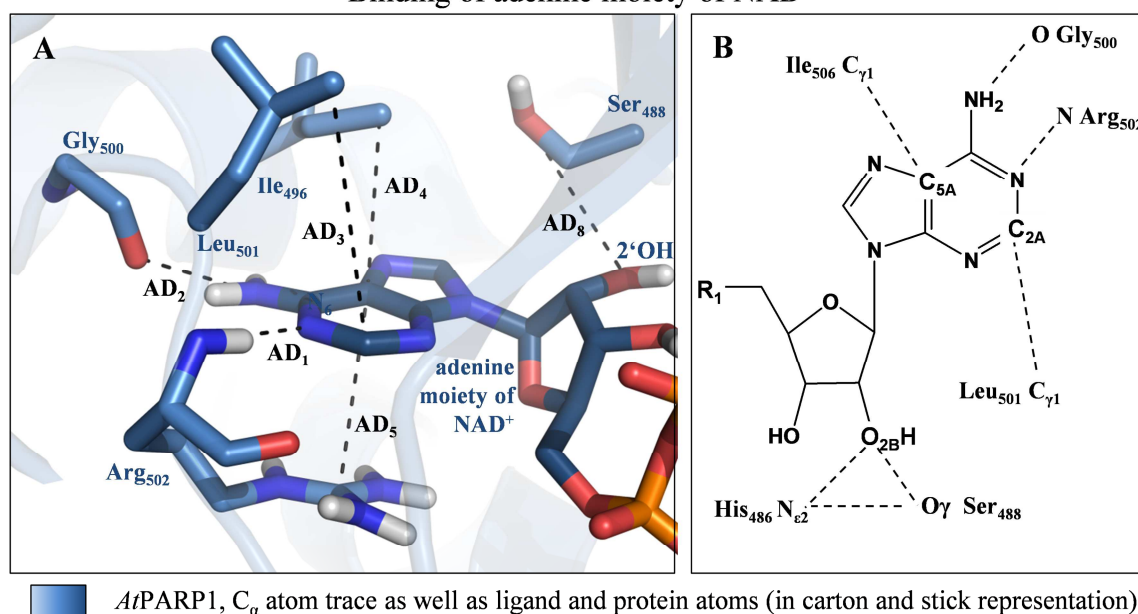


Figure 3.9: Positioning of the adenine moiety of NAD⁺ in AtPARP1 view into active site of AtPARP1; non-polar hydrogens omitted; dotted lines and AD₁ - AD₈ represent interactions or contacts described in Table 3.8

3.4.4 The role of the catalytic glutamate

The conserved catalytic glutamate Glu₆₁₄ (AtPARP1 numbering) is reported to be essential for ADP-ribosylation (e.g. Glu₉₈₈ in HsPARP1) in animal PARP.¹⁷ The E₉₈₈K mutation in HsPARP1 leads to a 98.5% loss of enzyme activity¹⁶ because, the oxygen atoms O_{ε1} and O_{ε2} of the side chain adjust the ribose units of the donor and acceptor structures before the catalytic reaction can take place (1.1.3). While the catalytic reaction itself cannot be captured with methods of molecular mechanics, the hypothesis was that the positioning and stabilisation of the ribose units through Glu₆₁₄ should be detectable in MD simulations of AtPARP1, including the substrates NAD⁺ and CNA.

In the Md simulations, the $O_{\epsilon 1}$ of Glu₆₁₄ forms a stable hydrogen bond to the O_{2D} of NAD⁺, another hydrogen bond exists between the $O_{\epsilon 2}$ of Glu₆₁₄ and the O_{3B} of CNA (Table 3.9 and Figure 3.10, interactions Glu₁ and Glu₂). The adjustment of the two riboses can be evaluated by measuring the distance between the attacking oxygen of the 2' hydroxyl of CNA and the carbon atom that forms the oxacarbenium during the catalytic reaction. The mean distances between these two atoms during the equilibrated simulation times were 3.48, 3.71 and 3.56 Å in the MD simulations with only moderate deviations (Table 3.9 and Figure 3.10, interaction Glu₃).

Analysis of MD simulations of the catalytic domain of *At*PARP1

The role of the catalytic glutamate 614

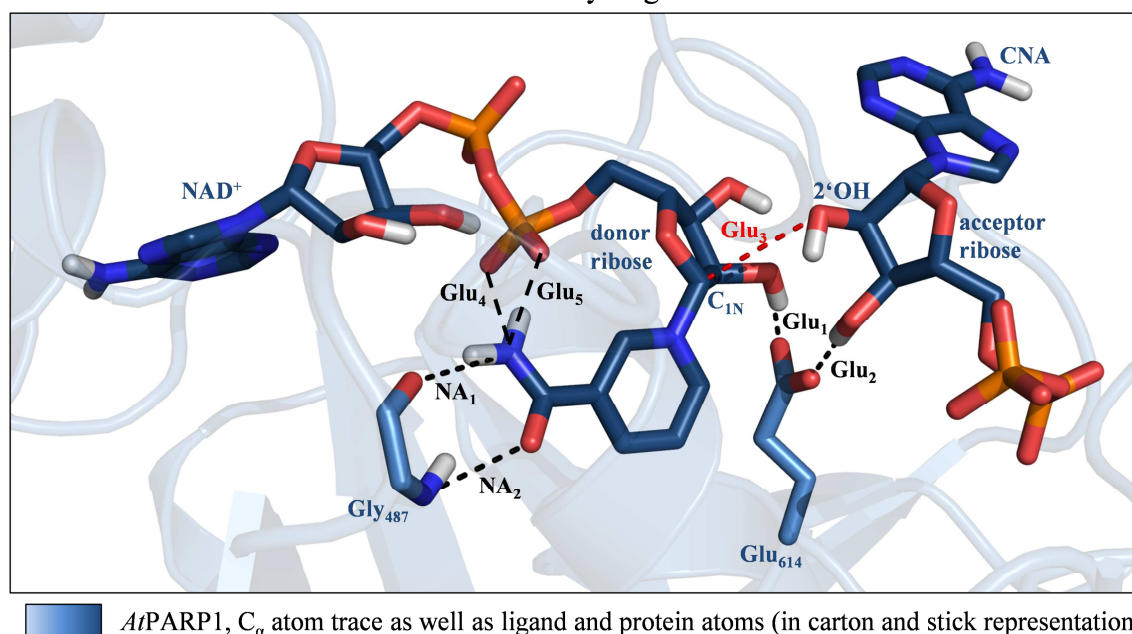


Figure 3.10: Positioning of the nicotinamide moiety of NAD⁺ in *At*PARP1 view into active site of *At*PARP1; non-polar hydrogens omitted; dotted lines and NA₁ - NA₂, as well as Glu₁ - Glu₃ represent interactions or contacts described in Table 3.7. and Table 3.9

In the modelled structure of NAD⁺ in *Gg*PARP1, the $\epsilon 2$ oxygen of Glu₉₈₈ (*Gg*PARP1 numbering) showed a distance to the 2'OH of nicotinamide ribose of 2.8 Å, which is in accordance to the observed distances in the simulations found between the $\epsilon 1$ of Glu₆₁₄ (*At*PARP1 numbering) and the 2'OH of nicotinamide.⁷² The distances between the $\epsilon 2$ oxygen of Glu₆₁₄ (*At*PARP1 numbering) and the 2'OH ribose of CNA confirm the results obtained from the modelling of NAD⁺ in *Gg*PARP1 where a distance of 2.7 Å was reported.³¹

Finally, an intramolecular hydrogen bond between the nicotinamide amide nitrogen N_{7N} and the O_{1N} (or O_{2N}) of the diphosphate moiety restrains the conformation of NAD⁺. This conformational restriction promotes the position of the nicotinamide ribose with respect to the

acceptor ribose and therefore positively influences the hydrogen bonding pattern of Glu₆₁₄. This intramolecular hydrogen bond is present in the majority of the simulation time (interactions Glu₄, Glu₅, Glu_{4&5} in Table 3.9 and Figure 3.10). This observation is in accordance to the crystal structure of Diphtheria Toxin where the distance between N₇N and O₂N of NAD is between 2.9 and 3.2 Å, indicating hydrogen bonding interactions.

Table 3.9: Results of MD simulations - comparison of experimental data III

interacting atoms			MD simulation analysis		
			distance (Å) or h-bond frequency		
			297.9999 K	298.0000 K	298.0001 K
Glu _{1_d}	NAD O ₂ D	Glu ₆₁₄ O _{e1}	2.59 ± 0.09	2.59 ± 0.10	2.60 ± 0.09
Glu _{1_h}	NAD O ₂ D	Glu ₆₁₄ O _{e1}	99.9 %	99.9 %	100 %
Glu _{2_d}	CNA O ₃ B	Glu ₆₁₄ O _{e2}	2.60 ± 0.10	2.60 ± 0.10	2.59 ± 0.09
Glu _{2_h}	CNA O ₃ B	Glu ₆₁₄ O _{e2}	99.9 %	99.8 %	99.9 %
Glu ₃	NAD C ₁ D	CNA O ₂ B	3.48 ± 0.21	3.71 ± 0.33	3.56 ± 0.22
Glu ₄	NAD N ₇ N	NAD O ₁ N	99.7 %	71.9 %	58.5 %
Glu ₅	NAD N ₇ N	NAD O ₂ N	0.0 %	0.0 %	28.3 %
Glu _{4 and 5}	NAD N ₇ N	NAD O _{1/2} N	99.7 %	71.9 %	86.8 %
Glu ₆	CNA O ₂ B	Tyr ₅₃₁ O _η	8.96 %	17.9 %	33.1 %

In the crystal structure of GgPARP1 (PDB code 1A26), there are formed two hydrogen bonds between the acceptor ribose and the catalytic glutamate 988. Simultaneously, another hydrogen bond between the 2'OH of the acceptor ribose and a water molecule (Wat₃₇ in 1A26, Figure 3.11) is present.³¹ In the same publication, the donor substrate NAD⁺ was modelled into the active site of the crystal structure. Superposition of the crystal structure and the NAD⁺-containing model revealed that the Wat₃₇ oxygen atom superimposes with the donor ribose carbon C_{1N}, indicating that Wat₃₇ in 1A26 mimicks the electrophilic C₁ of the donor ribose (1.1.3).³¹ Both observations led the authors to suggest that Glu₉₈₈ directly increases the nucleophilicity of the 2' oxygen of the acceptor ribose, while adjusting its position in favour of a nucleophilic attack. This is shown in the right panel of Figure 3.11.

An analogue view into the active site of AtPARP1 (left panel of Figure 3.11) reveals a similar picture. The acceptor and donor substrates CNA and NAD⁺ are hydrogen-bonded to Glu₆₁₄, and therefore adjusted for initiating the catalytic reaction. Superposing the AtPARP1 model and 1A26 results in a distance of only 0.9 Å between the C_{1N} carbon of the donor ribose and Wat₃₇ in 1A26, confirming the statement of Wat₃₇ as a C₁-mimicking atom in 1A26. It also confirms the quality of the homology model of AtPARP1 since the 2'OH is in favourate

position and distance for a nucleophilic attack (as shown in Table 3.9 and Figure 3.11). But since Glu₆₁₄ does not form a hydrogen bond to the 2'OH of CNA, the nucleophilicity of the attacking oxygen is not increased by the Glu₆₁₄. Instead, as indicated by interaction Glu₆ in Table 3.9, a hydrogen bond between the 2'OH and the hydroxyl group of Tyr₅₃₁ in *At*PARP1 is observed, suggesting that the polarisation of the attacking oxygen might also be possible through a nearby tyrosine O_η. In *Gg*PARP1, it is confirmed that the mutation Y₉₀₇N decreases activity to 1.1 % in relation to the wild-type. An involvement of this tyrosine in activating the catalytic reaction would augment the importance of this tyrosine (e.g. Tyr₅₃₁ in *At*PARP1 and Tyr₉₀₇ in *Hs*PARP1) such that its side chain does not only stabilise the nicotinamide moiety of the donor, but is also involved in the catalytic reaction itself.

Analysis of MD simulations of the catalytic domain of *At*PARP1

The role of the catalytic glutamate 614, II

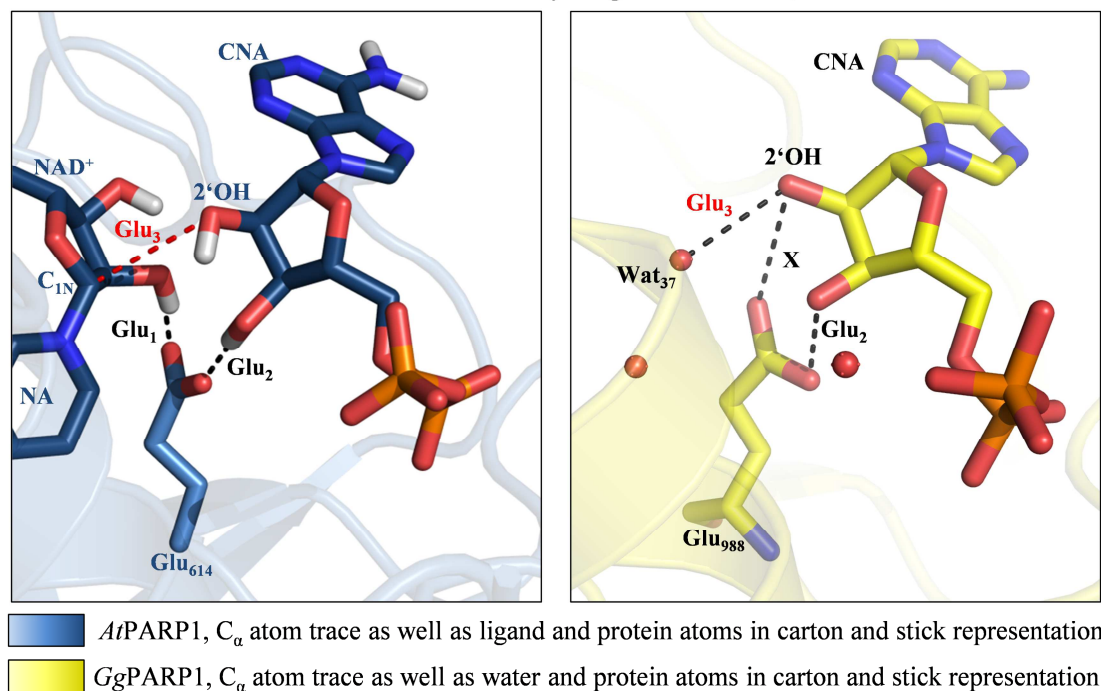


Figure 3.11: The role of the catalytic glutamate in the catalytic reaction view into active site of *At*PARP1; non-polar hydrogens omitted; dotted lines and Glu₁ – Glu₃ represent interactions or contacts described in Table 3.7 and Table 3.9, interaction X only present in PDB entry 1A26, position of Water 37 in PDB entry 1A26 equivalent to position of C_{1N} of the donor ribose in *At*PARP1

It might also be that the adjustment of the acceptor ribose through Glu₉₈₈ in PDB entry 1A26 is only observed when no donor but only an acceptor substrate is present (as in PDB entry 1A26, Figure 3.12) and the hydrogen bond and activation pattern changes upon the presence of both the donor and acceptor substrate. Crystal structures having both substrates or a transition state-analogue present in the active site would gain more insights of the catalytic reaction.

3.4.5 Positioning of the adenine moiety of the acceptor structure

In the crystal structure of *Gg*PARP1, the side chain of a conserved methionine residue (Met₈₉₀, *Gg*PARP1 numbering) was found to be positioned in parallel to the adenine ring of CNA, indicating that there are hydrophobic contacts between the protein and the acceptor ADP-ribose chain that stabilise the positioning of the latter one. To investigate whether analogue interactions occur in *At*PARP1, interactions between the modelled CNA and proximate amino acids in *At*PARP1 were analysed.

To investigate the relationship between Met₅₁₄ (representing the analogue of Met₉₈₀ in *Gg*PARP1) and CNA in *At*PARP1, the distance M_1 between the S_δ atom of Met₅₁₄ and the C_{8A} atom of the adenine moiety of CNA was monitored during MD simulations (interaction M_1 in Figure 3.12). Since the adenine ribose of CNA was kept in position via a stable hydrogen bond between Glu₉₈₈ and the 2'OH of the adenine ribose in *Gg*PARP1 (3.4.4), displacements of the adenine moiety were analysed by measuring the angle of torsion (interaction TA in Figure 3.12 and Table 3.10) between the atoms C_{2B}, C_{1B}, N_{9A} and C_{4A} of CNA (shown as atoms 1, 2, 3 and 4 in Figure 3.12).

Analysis of MD simulations of the catalytic domain of *At*PARP1

Binding of adenine moiety of CNA

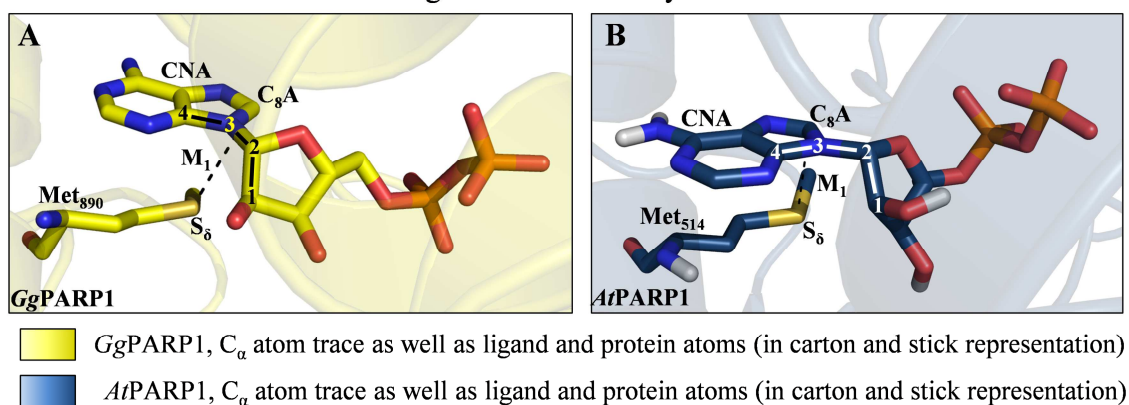


Figure 3.12: Positioning adenine moiety of CNA in *Gg*PARP1 and *At*PARP1
Views into active sites of *Gg*PARP1 (A) and *At*PARP1 (B); non-polar hydrogens omitted; dotted lines M_1 represents the distance between S_δ of methionine residue (Met₈₉₀ and Met₅₁₄, *Gg*PARP1 and *At*PARP1 numbering, respectively) and C_{8A} of CNA. The angle of torsion, TA, is represented by atoms 1, 2, 3 and 4

In only one of three MD simulations (run at 298.0001 K), analogue observations were found in comparison to *Gg*PARP1, while large deviations from *Gg*PARP1 results were found in the other two MD simulations (interactions M_1 and TA in Table 3.10 and Figure 3.13). Analysing additional data, such as the hydrogen bond pattern of the N_{6A} atom of CNA with the protein

(interactions $CNA_1 - CNA_3$ in Table 3.10) indicates that time-dependent conformational changes of CNA might have influenced those 5 parameters shown in Table 3.10.

Table 3.10: Results of MD simulations - comparison of experimental data IV

interacting atoms in			MD simulation analysis			Experimental results
CNA	<i>At</i> PARP1		distance (Å), h-bond frequency or angle (°)			
			297.9999 K	298.0000 K	298.0001 K	<i>Gg</i> PARP
M_1	C_8A	$Met_{514}S_8$	4.35 ± 0.84	4.90 ± 0.92	3.65 ± 0.27	3.57 \AA
TA	C_2B, C_1B, N_9A, C_4A (in CNA)		156.7 ± 99.8	221.4 ± 51.2	62.7 ± 9.9	82.1°
CNA_1	N_6A	$Gly_{512}O$	6.4 %	75.1 %	0.5 %	---
CNA_2	N_6A	$Thr_{511}O_{81}$	18.2 %	24.9 %	46.9 %	---
CNA_3	neither CNA_1 nor CNA_2		75.4 %	0.0 %	52.6 %	observed

Analysis of MD simulations of the catalytic domain of *At*PARP1

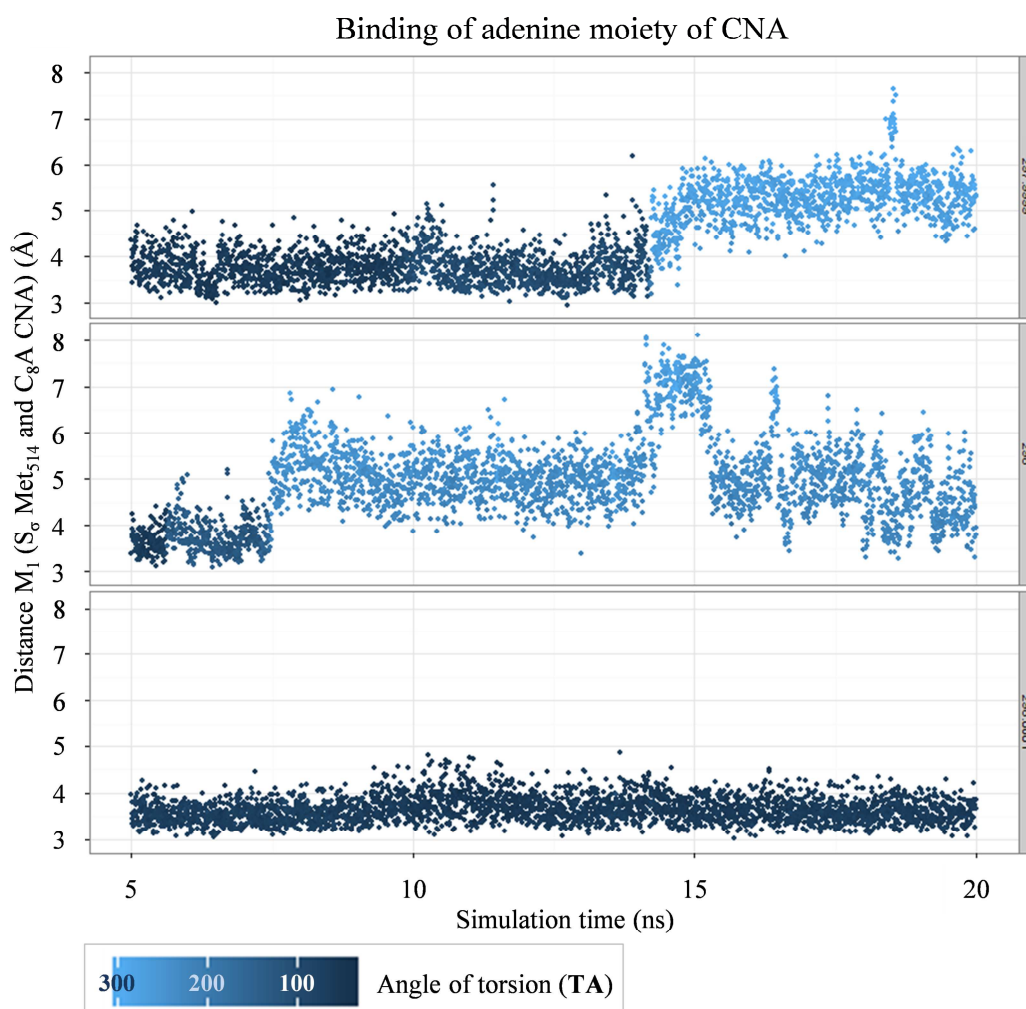


Figure 3.13: *At*PARP1 MD simulation analysis, adenine moiety of CNA positioning I
Time-dependent distances M_1 in relation to angle of torsion TA. Low values of M_1 (around 3.6 \AA) are associated with TA values below 80° , which is in accordance with observed data in *Gg*PARP1.

To determine the relationship between the fluctuations of the Met_{514} side chain and CNA,

interactions M_1 and TA were analysed in a time-dependent manner (Figure 3.13). The analysis shows strong evidence that low distances of M_1 (around 3.8 Å) are associated with torsion angles TA similar to those found in GgPARP1 (interaction TA in Table 3.10 and Figure 3.13, dark blue points at simulation run at 298.0001 K).

For identifying associations between the hydrogen bonding pattern and M_1 , as well as TA (representing all 5 interactions listed in Table 3.10) in a time-dependent manner, those parameters were analysed (Figure 3.14, Figure 5.12, Figure 5.13, and Figure 5.14).

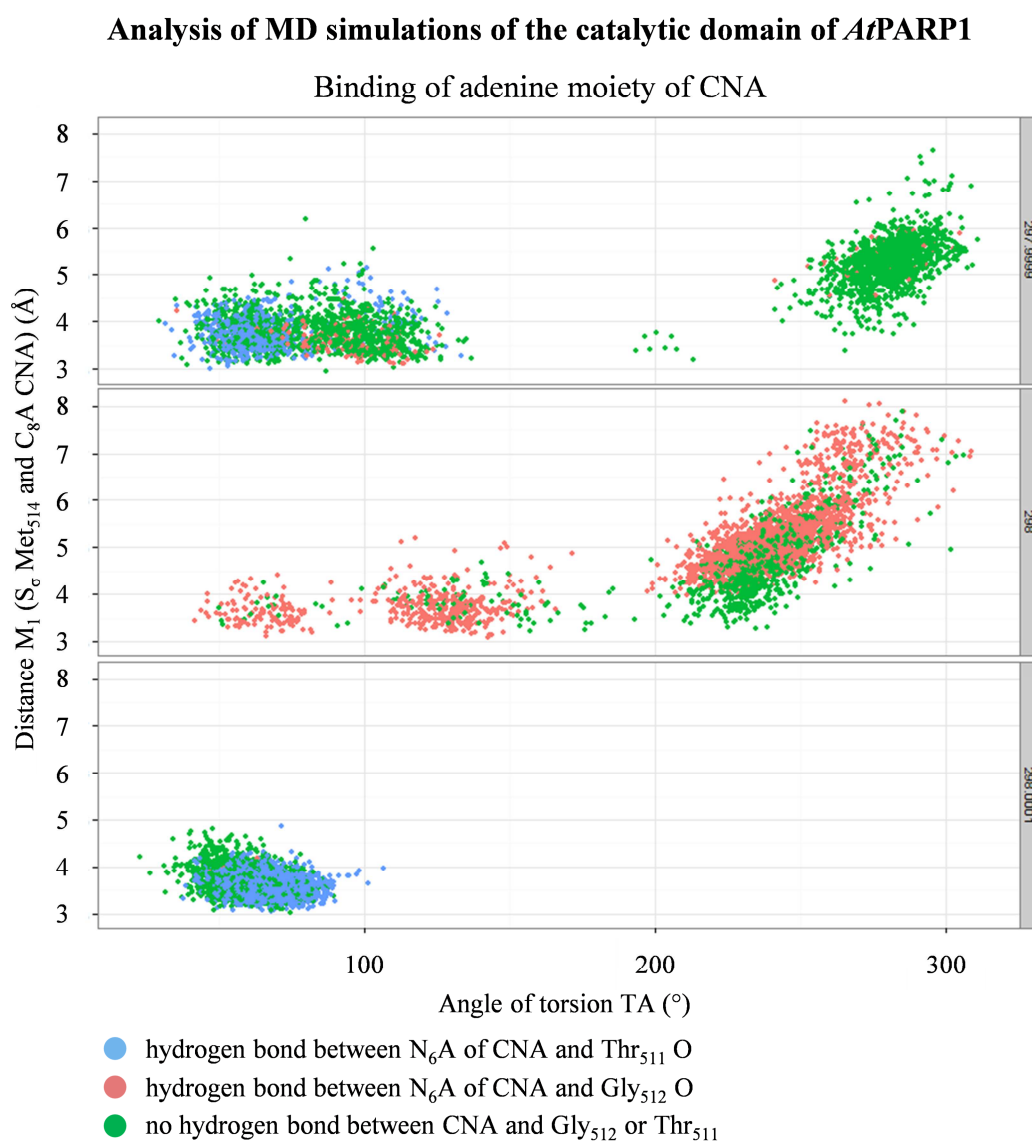


Figure 3.14: *At*PARP1 MD simulation analysis, adenine moiety of CNA positioning II
Hydrogen bonding pattern of CNA as a function of the angle of torsion (TA) and distance M_1 .

From Figure 3.14, Figure 5.12, Figure 5.13, and Figure 5.14, one can confirm the association between TA and M_1 , but also the corresponding hydrogen pattern. There is strong evidence that certain values of M_1 and TA are clustered and each cluster represents a certain hydrogen

bonding preference (Figure 3.14). There is also the picture emerging that the adenine moiety of CNA is very flexible since the variations in the angle of torsion (TA) can result in a flip of the adenine moiety (as indicated by comparison of the crystal structure data with a snapshot of MS simulation in *At*PARP1 in Figure 5.12). Despite the flexibility of CNA, the acceptor structure is still held in position via strong hydrogen bonds, e.g. to Glu₆₁₄ (3.4.4).

Additional hydrogen bonds with Gly₅₁₂ and Thr₅₁₁ support the fixation of CNA. But there is evidence that these hydrogen bonds have little influence on CNA binding. First evidence is that favourable values of M_1 and TA alone can result in a positioning (Figure 3.13 and Figure 5.14) being close to the positioning observed in *Gg*PARP1 (Figure 3.12, A). Secondly, only CNA that is not hydrogen-bonded to backbone atoms of Thr₅₁₁ or Gly₅₁₂ coincides with data in *Gg*PARP1 where CNA is not hydrogen-bonded to corresponding amino acids. This supports the assumption that not the hydrogen bonding of the adenine moiety of CNA, but the hydrogen bonding of the ribose moiety of CNA, as well as hydrophobic interactions are responsible for CNA stabilisation.

Analysing the MD simulations in the context of the acceptor structure CNA, one can conclude that the MD simulations are able to capture the interactions observed in the crystal structure of *Gg*PARP1.

The active site of *At*PARP1, as well as the donor and acceptor substrate positioning, which can explain recognition and binding of the substrates, could be modelled in sufficient quality. Therefore the quality of the homology model is of adequate quality to be used as a model for virtual screening and docking analysis.

3.5 Docking

3.5.1 Docking program selection

To select the suitable docking program for virtual screening, it was investigated, which docking program could place the set of 142 known human PARP1 inhibitors (2.1.3) in a correct pose (2.6.5) into the active site of the *Hs*PARP1 model. It was also assessed if NAD⁺ (2.1.1) could be docked correctly. Descriptions of docking parameters are explained in 2.6.3 and 2.6.4. Results of docking trials are summarised in Figure 3.15 and 5.6.1.

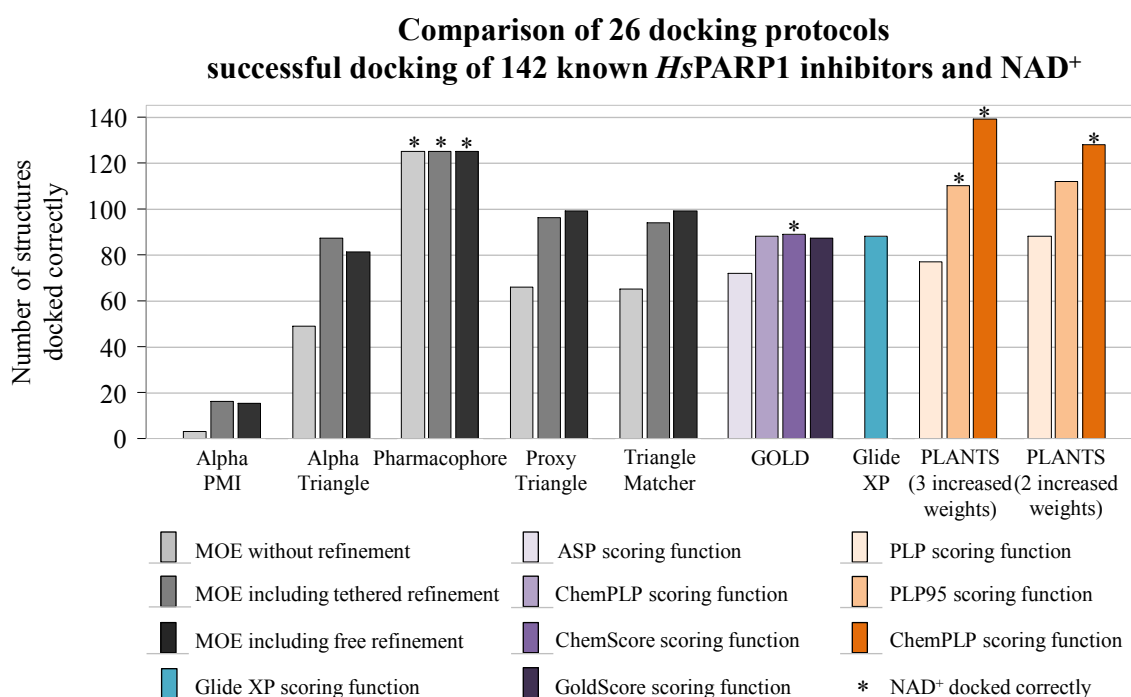


Figure 3.15: Comparison of docking programs

grey: MOE scoring functions including either no, tethered or free refinement of each pose. blue: Glide XP; violet: GOLD scoring functions ASP, ChemPLP, ChemScore and GoldScore. orange: PLANTS scoring functions PLP, PLP95 and ChemPLP; * indicates successful docking of NAD⁺ in a correct pose (2.6.5).

Among the different docking programs giving 26 different docking protocols, 7 of them were able to dock NAD⁺ into *Hs*PARP1's active site. Also, PLANTS's ChemPLP and MOE's pharmacophore docking parameterisations resulted in more than 80% of the Novikov inhibitor data set docked correctly into *Hs*PARP1's active site. Among those, PLANTS's ChemPLP scoring function outperforms all other parameterisations because it docked 139 *Hs*PARP1 inhibitors, as well as NAD⁺, correctly into *Hs*PARP1's active site. Since the PLANTS protocol I with three increased hydrogen bond weights outperforms all other docking programs, a second protocol, PLANTS protocol II, where only two hydrogen bonds were increased, was included into the analysis. Even in the PLANTS protocol II, ChemPLP's

performance was better than all other docking programs. Based on these results, the docking program PLANTS together with scoring function ChemPLP was chosen for further investigations of differentiation of known *Hs*PARP1 inhibitor from decoy structures and for the virtual screening process.

3.5.2 Receiver Operator Characteristics (ROC) curve

To investigate if the PLANTS protocols I and II, both with ChemPLP scoring function, are able to discriminate true ligands from decoy structures, the docking performance of both protocols was assessed with Receiver Operation Characteristics (ROC) curves. For both protocols, the data sets of *Hs*PARP1 inhibitors (2.1.3) and *Hs*PARP1 decoys (2.1.4) were docked into the active site of *Hs*PARP1 in 10 independent docking runs. The ROC curves were plotted and AUC calculated for PLANTS docking score TOTAL SCORE. The mean AUC for protocol I was 0.861 ± 0.005 , whereas the mean AUC for the protocol II, the mean AUC was 0.879 ± 0.010 (mean \pm 95% CI). Two sided unpaired *t*-tests at 5% significance level α showed a significant difference between both mean AUCs ($P < 0.001$).

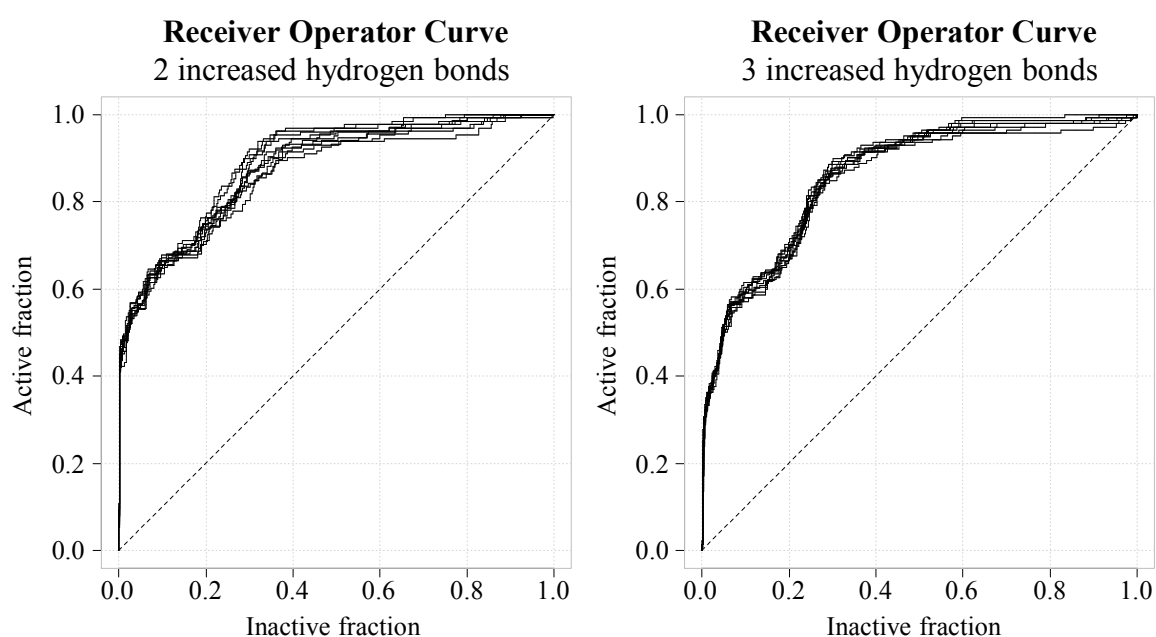


Figure 3.16: ROC curves of PLANTS docking protocols I and II

Both protocols were sufficiently able to discriminate, and there was a significant difference in the detectable performance, suggesting that protocol II was more powerful in discrimination. Increasing only two instead of three hydrogen bond weights and therefore lowering power of the guidance of the docking poses did affect the discrimination quality positively. Visual

inspection of one hydrogen bond contribution score had the same effect as steering the hydrogen bond formation through increased weights. The results are displayed in Figure 3.16.

The directory of useful decoys (DUD) contains inhibitors and decoy structures for each target in a ratio of about 1:40. The DUD ligand set for *Hs*PARP1 contains 35 PARP inhibitors. Among those, one can find structures sharing phthalazinone structures, as well as the *Hs*PARP1 inhibitors 4ANI and 3AB. The majority of structures is based on the same scaffold. Therefore, DUD's PARP ligand set covers the chemical space of human PARP inhibitors not sufficiently. The data set published by Novikov contains a variety of structural classes. They are collected from 6 different publications which cover more chemical classes and more derivatives per class. The Novikov data set covers the chemical space of *Hs*PARP inhibitors much better than the DUD ligand set. Furthermore, the data set by Novikov is approximately four times as large as the DUD ligand set. Together with the wide range of structural diversity it represents a much better random sample of human PARP inhibitors. The standard inhibitor and decoy ratio of about 1:40 is increased to about 1:9.5. Since no EF studies, but only ROC plots for discrimination studies are used, this increase of ratio has no negative influence on the performance of the studies.

Although the Novikov data set can be assumed to be a SRS, this assumption cannot be proven, since it is impossible to know all potential PARP inhibitors. If the whole chemical space of human PARP inhibitors would be discovered, there would not be any need for further development or improvement of PARP inhibitors. Also, the assumption that DUD decoys, although being more than 1300 structures, would cover the chemical space of potential PARP decoys, is not valid for the same reason. Furthermore, among all decoy structures there might be some so called false false positives. Decoy structures are meant to be not active although having similar properties and shape. But they have not been *in vitro* verified as PARP decoys. In ROC analysis some of them will be false positives because they are classified as active although they should not bind to PARP. But since the decoys are not tested to be real decoys, some of the false positives will be active against PARP which are then called false false positive outcomes.

Up to now, there is no ideal data set (for any target) that perfectly fits all assumptions. But the DUD decoy set is the only published decoy set for PARP. The Novikov PARP ligand data set was the most diverse set. Both sets therefore represent appropriate data sets for investigation of docking performance.

3.5.3 Inference for data set docking score distributions

For each of the 10 docking runs i , the corresponding means \bar{x}_i and standard deviations s_i of the docking score distributions of decoy and ligand samples were calculated. The pooled mean \bar{x}_{pooled} and pooled standard deviation \bar{s}_{pooled} of the 10 independent docking run means \bar{x}_i and standard deviations s_i of *Hs*PARP1 inhibitors and *Hs*PARP1 decoys were used as estimates for the populations of PARP ligand and PARP decoy docking scores. These estimates ($\bar{x}_{pooled} \pm \bar{s}_{pooled}$) are (-153.00 ± 16.76) for the Novikov data set and (-123.32 ± 19.54) for *Hs*PARP1 decoy data set.

3.5.4 Normal approximation of docking score distributions

The assumption of normality was checked by overlaying the histograms of the ligand and decoy samples with a corresponding normal distribution $N(\bar{x}_i, s_i)$ which was estimated from the sample mean \bar{x}_i and the sample standard deviation s_i of each docking run i . Exemplary, the overlay of the histogram and the estimated normal distribution $N(\bar{x}_i, s_i)$ for the 3rd docking run of ligands and decoys are shown in Figure 3.17

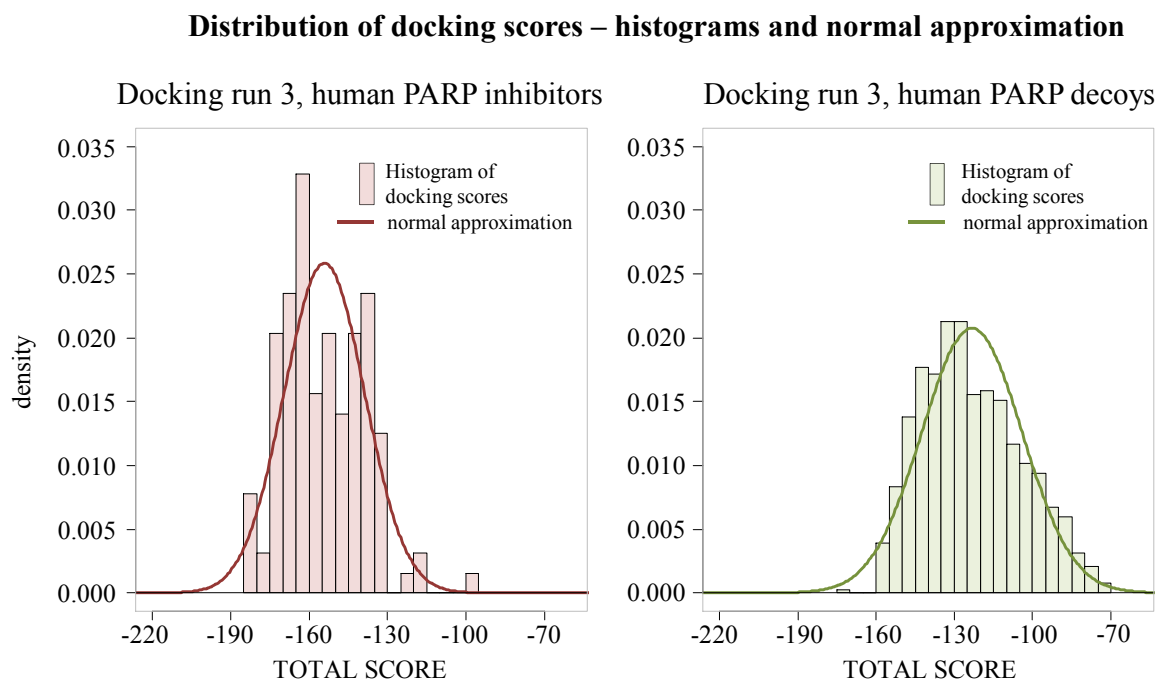


Figure 3.17: Docking score distributions – normal approximation I
Normal approximation of the docking scores (displayed as histograms) was done using the pooled means and pooled standard deviations described in 3.5.3

For both the ligand and decoy data, the normal approximations do not perfectly match the actual docking score distribution. To assess the difference between the observed docking

scores and their normal approximations, the cumulative distribution functions (cdf) of each docking run i , CDF_i , and corresponding normal distribution $N(\bar{x}_i, s_i)$ were compared. This was done for both the ligand and decoy samples. In the cases of inhibitors and decoys, the CDF display a similar, but not identical, shape. The difference between the sample CDF_i and the CDF_i of the estimated normal distribution $N(\bar{x}_i, s_i)$ equals the error of both CDF and is called e_{CDF_i} . Together with the CDFs, the e_{CDF_i} was investigated and plotted. The observed and approximated cdf for the 3rd docking run for ligand and decoy data set are shown in Figure 3.18, A and C. The corresponding errors between observed and approximated data (e_{CDF}) are shown in Figure 3.18, B and C, respectively.

Docking run 3, normal approximation and corresponding errors

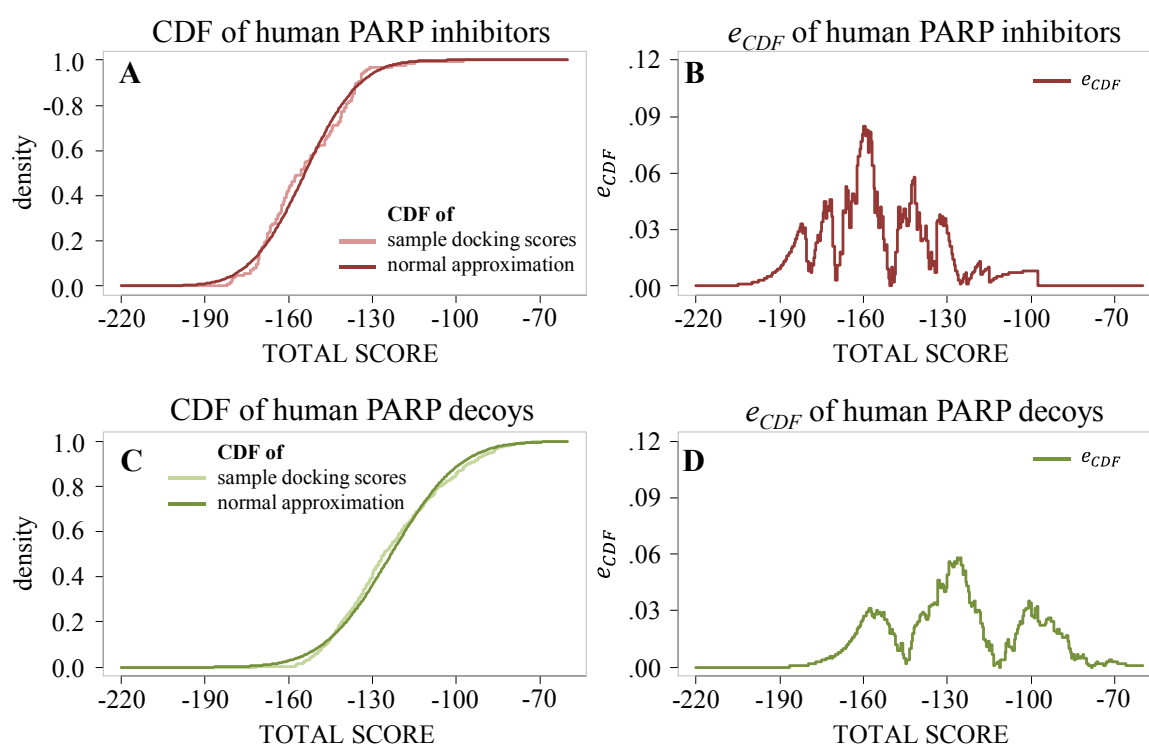


Figure 3.18: Docking score distributions – normal approximation II
comparisons of CDF for HsPARP1 inhibitors (A) and HsPARP1 decoys (C), and corresponding errors of CDF from normal approximation for HsPARP1 inhibitors (B) and HsPARP1 decoys (D),

For the ligand data set, the errors, except for one peak around docking scores of -160, are below 6%. In the case of the decoy data, the errors are below 6% for the complete range of docking scores, only having three peaks. This indicates that, although the normal approximations do not perfectly match the observed data, the corresponding errors are small enough, so that the approximations can be used instead of the observed data. The equivalent examination of errors for all 10 docking runs i is shown in Figure 5.8

3.5.5 HsPARP1 inhibitor docking score threshold derivation

The normal approximations of docking score distributions for HsPARP1 decoys $N_{decoys}(-123.32,19.54)$ and HsPARP1 inhibitors $N_{inhib}(-153.00,16.76)$ were used to derive a docking score threshold upon which new structures should be classified as active or inactive. (5.6.2) Since both distributions overlap, false positive and false negative outcomes are inevitable. The probabilities of committing type I and type II errors were assessed using power analysis. Power analysis was performed by setting the rate of committing a type II error to 5% which equals a statistical power 95% (2.7.1). The fixed type II error rate at the 5% level corresponds to the 5% percentile of the HsPARP1 decoys distribution. The corresponding TOTAL SCORE in PLANTS is -155.46. This score was set as a new docking score threshold for discrimination between human PARP1 inhibitors and human PARP1 decoys. The corresponding type I error probability for this docking score is 55.84%

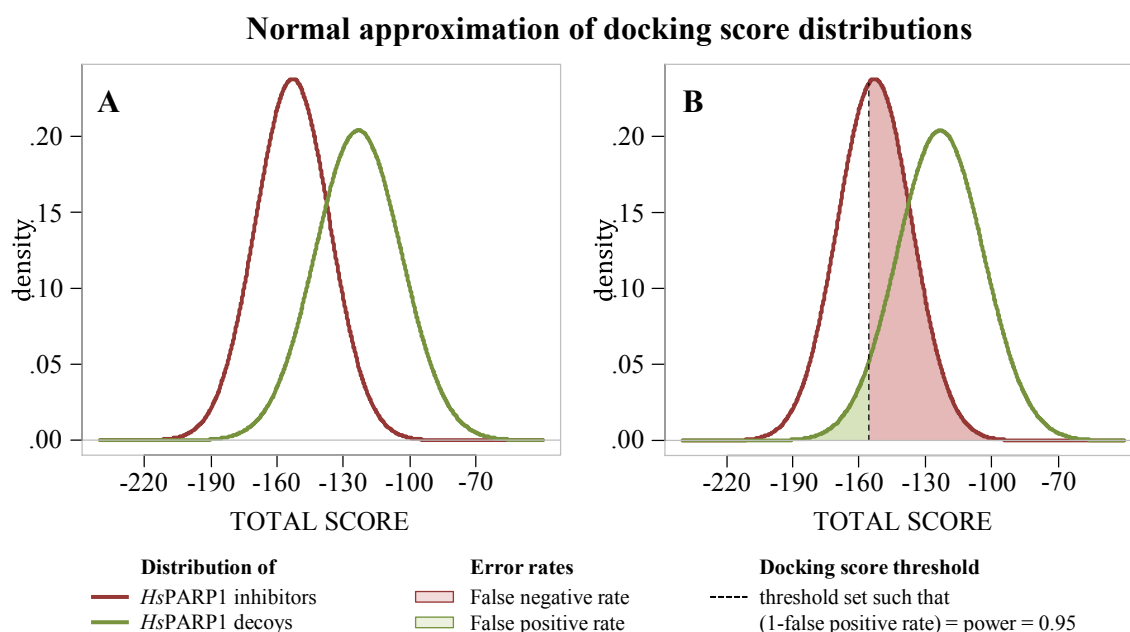


Figure 3.19: Normal approximation III

Based on defined docking score distributions (A), the docking score threshold (black dashed line in B) was set such that type II error rates are fixed at 5%. This threshold leads to a type I error rate of 55.84%. Both error rates are displayed as light red and light green areas in B.

Figure 3.19 A displays the approximated docking score distributions for the ligand (red line) and decoy (green line) data sets. The left side of Figure 3.19, B, displays the docking score of -155.46 as a black dashed line. This line corresponds to the type II error rate being displayed as a light green area, and the type I error rate as indicated as a light red area in Figure 3.19, B.

Based on power analysis and the established assumptions that the data sets, on which the power analysis was performed, contain representative samples of HsPARP1 ligands and

decoys and that their TOTAL SCORE distributions follow the normal distributions as described in 3.5.4, then the following can be stated:

If a commercial database contains potential *Hs*PARP1 inhibitors and structures that do not inhibit *Hs*PARP1 and if these two classes of structures follow the distributions as were inferred in 3.5.4 and if this database is screened with the developed PLANTS docking protocol II and ChemPLP scoring function, then potential *Hs*PARP1 inhibitors and decoys could be selected solely based on the docking score with the following characteristics: Among the selected structures that would be classified as inactive based on the docking score (e.g. have as TOTAL SCORE >-155.46), 95% of these would be correctly classified as inactive if they were tested *in vitro*. Among the selected structures that would be classified as active based on the docking score (e.g. have as TOTAL SCORE < -155.46), 44.16% (= 100% - 55.84 %) of them would be correctly classified as active if they were tested *in vitro*, too.

Table 3.11: Characteristics of the docking threshold

characteristic	Novikov ligands	DUD decoys
mean number of correctly docked structures	129	760
mean number of structures missing the threshold	64	14
mean type I error, based on power analysis	55.84 %	
mean type I error, observed	49.60 %	
difference	6.24 %	
mean type II error, based on power analysis		5.00 %
mean type II error, observed		1.88 %
difference		3.12 %

further details of the 10 docking runs are listed in 5.6.4, 5.6.5 and 5.6.6

Statistical power analysis can be used to define conditions which need to be fulfilled so that an existing effect can be correctly detected by the method investigated. Assessing the power of a method or study is essential since not assessing the power of a process can lead to underpowered or overpowered methods – conditions that are not desired. The more underpowered or overpowered a study or method gets the more inefficient it is. Underpowered studies are unable to detect an existing effect and unable to reject the null hypothesis. Results that lack significance because they are underpowered can lead to misinterpretation of results and therefore to wrong understanding of the problem that is investigated. Instead of interpreting non-significant results as no evidence of an effect, it is sometimes interpreted as evidence of no effect. In clinical trials, insufficient power in

treatment studies can also be problematic or even unethical if participants are exposed to inferior treatment.²³⁹

Overpowered studies on the contrary are prone to finding effects that are essentially without any meaning. Such situations can occur when the number of observations is far bigger than 1.000 or hypotheses are tested on large databases. The power of any statistical test affects statistical significance. In an extreme example it was shown that two identical groups have been found different with statistical significance ($p < 0.05$). Having one million observations, the actual SPSS-generated difference was -0.046 , which was not a meaningful, but still significant result.²⁴⁰ Therefore, finding an adequate level of power is a key point in study design. Initial experiments that give hints about the danger and costs in comparison to potential benefits (in any sense) are indispensable. A statistical power of 50% means 50 – 50 chances to reject the null hypothesis that was found to be false or being unable to detect a true effect in 50 percent of the time. Since statistical power is related to type I error, desired effect size and the number of observations, one has to adjust among all of these factors. Fixing statistical power and alpha levels result in an adjusted number of observations and effect size. False positive outcomes, the number of observations and different effect sizes can be translated into costs, and so are the benefits of high power. Therefore, it is always to ask: “Does the nature of the effect warrant the expense required to uncover it?”²⁴¹

The appropriate setting of statistical power ($1 - \beta$) is always problem-specific. There is no rule of thumb for choosing β . In 1988, Cohen²⁴² rationalised to set the power at 80% which can be translated into studies that have a probability of 20% to give a false positive result and a probability of 80% to correctly detect an existing effect. Cohen argued that α (typically set to 0.05) should be treated four times as serious as β (being set to 0.20) and balance the risks of committing type I and type II errors in that ways. From that time on, researches not only had to rely on Fisher’s 5% type I error criterion, but also on type II errors, Cohen’s recommendation became well-known as the five-twenty convention or the one-to-four rating of α and β errors.

The goal of power analysis of the data sets for human PARP1 ligands and human PARP decoys was to choose a docking score that identifies new structures (from commercial or inhouse databases) as potentially active or inactive, according to docking scores of contained structures. Therefore the assumption has to be made that structures in databases can also be classified in PARP inhibitors and PARP non-binders and that the docking score distribution

follows the ones that were calculated for the test sets of *Hs*PARP1 inhibitors and *Hs*PARP1 decoys. It is plausible that in commercial databases, the number of potential PARP1 inhibitors will be far smaller than the number of structures that do not inhibit the protein. The proportion of ligands and decoys in the test set was 142:1351 \approx 1:9. In the DUD, this proportion was set to approximately 1:40 for each target.¹⁵⁸

Considering type I errors or false negative docking outcomes, the generally used error rate $\alpha = .05$ would be inappropriate for practical reasons of *in vitro* testing of selected compounds. This will be explained on the 1st docking run in the *Hs*PARP1 ligand and decoys test set, where 128 of 142 inhibitors and 771 of 1351 decoys were docked correctly. The setting of $\alpha = .05$ equals keeping 95% of ligands for testing while dismissing 5% of the ligands. This setting would qualify 121 of 128 correctly docked ligands to be selected. The docking score threshold at which this amount of ligands would be classified as active would be -120.60. The number of decoy structures that pass this threshold is 432. This number corresponds to the number of false positives (and inactives) with according false positive rate (or β error) of 56.1% and a power or 43.9%. (Table 3.12, setting 1) Increased α -levels increase the docking threshold (e.g. making TOTAL SCORE more negative) which leads to less inhibitor and decoy structures to pass this criterion. Due to the increasing of the α -level and reduced number of selected decoys, the false positive rate decreases and statistical power rises (Table 3.12, setting 2). While the database-size approach neglects the docking score, in the example of the 1st *Hs*PARP1 docking with 899 structures, all nine selected would be inhibitors (Table 3.12, setting 2). Approaches having high statistical power have a significantly stricter docking threshold. Because of that, less structures from the pool of actives are selected which leads to an increased false negative rate (or α error). The stricter docking threshold is also the reason for a decreased number of inactives that will be selected and increases the number of inactives that will correctly be identified as such (Table 3.12, settings 4 and 5). This is the reason for increased statistical power.

Table 3.12: Compound selection based in different focal points

setting	focus	docking score threshold	number (and percentage) of selected ...				power (1- β)	
			inhibitors <i>n</i> = 128		decoys <i>n</i> = 771			structures total
1	$\alpha = .05$	-120.6	121	(95.0)	432	(56.1)	553	43.9 %
2	$\alpha = .10$	-127.6	115	(90.0)	322	(41.8)	427	58.2 %
3	1 % db size		9	(7.0)	0	(0.0)	9	100 %
4	$\beta = .90$	-148.4	73	(57.7)	77	(10.0)	150	90 %
5	$\beta = .95$	-155.4	55	(41.2)	38	(5.0)	93	95 %

Sticking to low α -levels would result in high proportions of inhibitors contained in the test set, but the number of decoy structures that pass the threshold according to these α -levels sum to a large amount of compounds that would have been bought based on the α -decision. The theoretical or statistical significance does not account for costs and benefits. These characteristics are assessed in evaluation of practical significance:

In contrast to statistical significance, practical significance measures the impact of real-world application of this docking threshold. In particular, practical significance could be defined as the benefits of an agrochemical company of having identified a new lead compound that increases abiotic stress tolerance in crop plants at a defined level. Practical significance could also measure the costs necessary to identify a hit from virtual screening. It incorporates the questions: “What are the costs of identifying a compound that increases stress tolerance to a certain level?” or “If the amount of financial support is limited to X, how many compounds can be tested if a single test costs Y €”. These questions require an analysis of how well the employed virtual screening is able to identify a hit or potential lead. At this point, statistical power analysis that focusses on specificity (or β errors) helps to answer this question since it estimates the number of compounds needed to screen to identify a hit. From an economical point of view it supports the decision how many compounds need to be tested (or how many money is spent on buying and testing) if there is a certain ratio of false positive and false negative outcomes contained in the sample to be screened. In Table 3.13, five different screening scenarios (settings) are compared. Since each screening scenario focusses on a different type of error, the number of structures that is tested is different in each setting.

If the costs of *in vitro* testing are assumed to be fixed at a level of 30 € per compound (neglecting personnel costs and overhead), than the following economic consequences arise that are given in Table 3.13:

Table 3.13: Statistical vs. practical significance

setting	focus	Practical significance / costs of measuring (in €)			
		total costs	identification of decoys	identification of hits	ratio of hit identification
1	$\alpha = .05$	16590	12960	3630	21.88 %
2	$\alpha = .10$	13110	9660	3450	26.31 %
3	1 % db size	270	0	270	100.0 %
4	$\beta = .90$	4500	2310	2190	48.67 %
5	$\beta = .95$	2790	1140	1650	59.14 %

Costs of measuring potential PARP inhibitors in relation to hit rates for different virtual screening settings at an assumed cost of 30 € per compound in the assay

In settings 1 and 2 which are focusing on α errors or on reducing false negative outcomes the number of selected compounds is high in comparison to β -driven settings 4 and 5 and so are the total costs of compound measuring. Among the high numbers of selected compounds the actual benefits (in identification of hits) are low in comparison to approaches 4 and 5. The greatest difference between the two approaches is occurring between settings 1 and 5. In (β -focused) setting 5, not only the total costs are less than 20% of those in setting 1, but the chance of identifying a hit in setting 5 is 270% of the chance in setting 1.

Setting 3 seems to be advantageous over all other settings at first sight. The costs are very low and the success rate is 100%. But the major drawback of setting 3 is that it selects only nine structures. All of those are positive but it is very likely that those structures share an already known core. The odds that these nine structures add knowledge to the problem under investigation are very low. Many VS strategies are based on already known structures or chemical classes and it is likely that those chemical classes are found in the first ranks of a ranked database. Furthermore, setting 3 completely ignores the docking score. Ignoring the docking score speeds up the whole virtual screening process since no docking score analysis has to be performed. But it is very likely that a follow-up VS has to be performed to run a more sophisticated VS run that also is able to identify more compound classes or searches the chemical space more rigorously.

These results reflect the advantages of power analysis. The number of compounds that might have no inhibitory effect on *At*PARP1 is reduced in large amounts while the percentage of active compounds is increased. As it is stated by Triballeau and coworkers, the β -focused strategy “may be advisable in small companies” and “is faster, cheaper, motivating, and apparently, the most efficient way to accelerate drug discovery”.²⁴³

Besides the advantages of deriving a threshold on specificity and β errors, there is one point that favours the focus on type I (or α) error and sensitivity: Selecting a less strict threshold using the classical one-to-four rating proposed by Cohen²⁴² dismisses less active structures. With the increase in the amount of selected active structures, the probability of selecting structures from diverse chemical classes raises.²⁴³ This can have tremendous effects on the study outcome since the broader the chemical space that is represented by the selected actives, the more knowledge can be gained about the target on which the compounds act or their mode of action.

While Cohen²⁴² suggested the $\alpha=0.05$ and $\beta=0.20$ convention for researchers that have no guidance how to choose α and β levels²⁴², this one-to-four rule has to be reconsidered in every occasion in which the risks and benefits of the test results can be estimated.²⁴³ As in the PARP virtual screening example, where β is decreased to 0.05, medical tests are designed in a way that the occurrence of type I errors is assumed to be less bad than type II errors, because wrongly detecting something on an actually healthy patient (type I error) and verifying later on (in follow-up experiments) that the first test was wrong is less harmful than telling an actually diseased patient that everything is well (type II error). In those cases, β is often chosen to be less than 0.005. Furthermore, purely focusing on α levels (together with a null hypothesis that assumes no effect) does not gain any knowledge about the investigated problem. Often, assessing α by testing against H_0 is meaningless, since it is already assumed that there is actually an effect. Taking the effect into consideration, one should always focus on β and statistical power.²⁴⁴

In 1933 Neyman and Pearson stated that there is no general rule for balancing type I and type II errors and that the leveling of power is problem-dependent has to be defined by the investigator.²⁴⁵ The power of the docking procedure was set to 95% which is a large deviation from Cohen's one-to-four rule (in which power is set to 80%). This high value of power was chosen because of its practical consequences. This is in agreement with the argument of Hubbard who states that this decision has "nothing to do with statistical theory but is based on context-dependent pragmatic considerations where informed personal judgment plays a vital role".²⁴⁶

If a database of 100,000 structures is screened and structures are selected based on the docking score a power of 95% results in correctly identifying 95% of all non-binders while retaining 5% for *in vitro* testing. The number of non-binders can be assumed to be much higher than the number of active structures (an optimistic example would be to have 95,000 inactives and 5,000 actives contained in the database). As a consequence, even a high power of 95% would allow $95,000 \times (1 - 0.95) = 4,750$ inactive structures to pass the filter. According to the one-to-four rule of 80% power, this number would be $95,000 \times (1 - 0.80) = 19,000$. Even under a high power of 95%, *in vitro*-measuring the activity of 4,750 inactive structures causes high costs but has little benefits and is highly time-consuming. Reducing the power of 80% would increase the costs by 400% without any beneficial increase (as indicated in Table 3.13). For this reason, a power of 95% might be even too low for large databases.

3.5.6 Derivation of docking score threshold for *At*PARP

3.5.6.1 Assumptions

The following assumptions had to be set prior to transfer the docking protocol for *Hs*PARP1 to *At*PARP. All concern the equality of the systems and their behaviour.

Assumption 1: The active sites of *Hs*PARP1 and *At*PARP are identical.

Explanation: Through superposition of active site residues of *Hs*PARP1 and *At*PARP it is clear that they are not identical since the RMSD values of active site residues are > 0 . But the RMSD in the active site region are low and amino acids in the active site, which are exchanged in both systems, have been set flexible during the docking and therefore the effect of differences is minimised.

Assumption 2: Both *Hs*PARP1 and *At*PARP are inhibited by the same inhibitors through the same mode of action and both *Hs*PARP1 and *At*PARP1 are not inhibited by the same decoy structures.

Explanation: This assumption cannot be proven because there is no estimate that states how well the chemical space for *Hs*PARP1 inhibitors is investigated. For *At*PARP, only three inhibitors are known (1.4.1).^{208,218} From a docking program's point of view, based only on atomic coordinates of the active site and assumption 1, this assumption can be assumed to hold true.

Assumption 3: The performance of *Hs*PARP1's docking procedure on *At*PARP is equal to the performance of *Hs*PARP1's docking procedure on *Hs*PARP1.

Explanation: This assumption can be tested by docking the data set of known *Hs*PARP1 ligands into the active site of *At*PARP. Assumption 3 can be used as hypothesis that can be tested, if assumptions 1 and 2 are assumed to hold. Docking scores for Novikov ligands were compared by unpaired two-sided Student's t-tests at a significance level of 0.005. The Null hypothesis in this test would be that there is no difference between *At*PARP and *Hs*PARP1.

3.5.6.2 Differences of the docking procedure between *Hs*PARP1 and *At*PARP1

Ten docking runs for *Hs*PARP1 were performed under equivalent conditions for *At*PARP1 and *At*PARP2 for the 142 structures containing Novikov data set (2.1.3). Ten conformations were produced per structure. For each docking run in *Hs*PARP1, *At*PARP1 and *At*PARP2 the number of structures that could be docked correctly according to 2.6.5 and the occurrence of

the correct pose were examined. The results are shown in Table 3.14 and Figure 5.9. Since in PLANTS, all poses are ranked by TOTAL SCORE in ascending order, it was also investigated, whether the correct pose could be found in pose number 1, which corresponds to the most negative TOTAL SCORE.

Table 3.14: Characteristics of docking procedure

characteristic	<i>Hs</i> PARP1	<i>Hs</i> PARP2	<i>At</i> PARP2
Mean of correctly docked ligands	129	123	111
Mean number of correctly docked ligands in pose 1	102	80	92
Mean percentage of correctly docked ligands in pose 1	79.5 %	65.5 %	83.3 %
Docking runs with 80 % > of correctly docked ligands	9	7	3
Docking runs with 95 % > of correctly docked ligands	5	4	0
Structures that could not be docked in any docking run	2	2	13
Structures docked correctly in all 10 docking runs	100	85	98

The mean number of structures that were docked correctly for *Hs*PARP1 is higher than the mean number of correctly docked structures for *At*PARP1 and *At*PARP2. In contrast to that, the number of structures that could be docked correctly in the first pose (and with most negative docking score) is highest for *At*PARP1.

The feature of docking procedure that was used for determination of a new docking score was the number of structures that could be docked correctly in each of the 10 docking runs. Table 3.14 shows that 100 of 142 structures were docked correctly in all 10 docking runs into *Hs*PARP1, while 85 and 98 structures were docked into *At*PARP1 and *At*PARP2 in all docking runs, respectively. These structures could be docked with confidence into the corresponding active sites and therefore were used for further analysis (Figure 3.20). The docking scores for all inhibitors, depending on whether they were docked with confidence into *Hs*PARP1 and *At*PARP2, are shown in Figure 5.9.

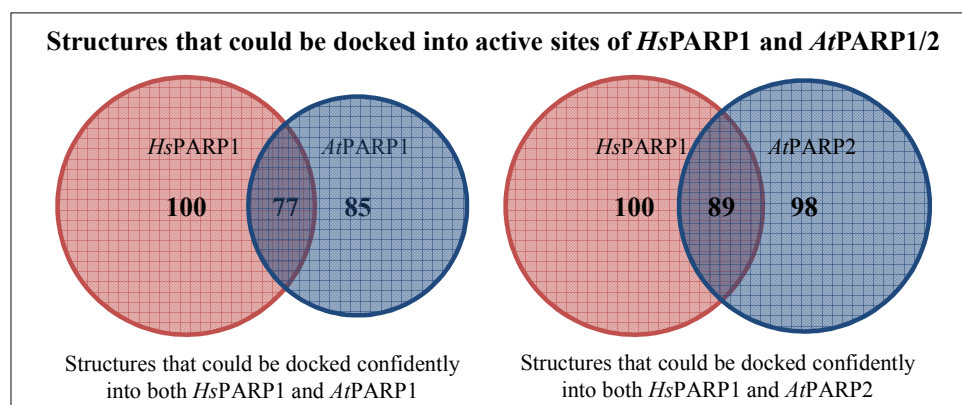


Figure 3.20: Number of structures being docked into *Hs*PARP1 and *At*PARP

The intersection of the sets of structures that were docked in 10 docking runs correctly contain most information for further analysis. The intersecting sets of structures that could be docked confidently into *Hs*PARP1 and either *At*PARP1 or *At*PARP2 contained 77 and 89 structures, respectively and is displayed in Figure 3.20.

Based on the 77 structures that were correctly docked into both active sites of *Hs*PARP1 and *At*PARP1 in all 10 docking runs, a new docking threshold was developed. The same procedure was performed for the 89 overlapping structures for *Hs*PARP1 and *At*PARP2. The differences of the mean docking scores were analysed with two-sided unpaired Student's *t*-tests at significance level $\alpha=0.005$ (2.7.3) which were performed for each of the 77 and 89 structures, respectively. The results of these tests are shown color-coded in Figure 3.21, together with the step of final derivation of new *At*PARP docking thresholds.

For the final derivation of a new docking threshold, only the ligands that fulfilled the prerequisites AND whose mean docking scores were significantly different were taken into account. For those remaining ligands, the median of docking score distributions of the differences were defined as the new docking threshold for *At*PARP1/2. This is illustrated in Figure 3.21, where the left panel (A, C) represents the derivation of the new dockings threshold for *At*PARP2, whereas the right panel (B, D) represents the derivation of the new docking threshold for *At*PARP1, respectively. The upper barplots (A, B) represent the mean docking score differences of the remaining ligands. The distribution of those differences are represented in the lower histograms (C, D). The medians of these histograms are represented by red lines, which implicate the median difference of docking scores.

The number of structures that were docked confidently into *At*PARP1 and *At*PARP2, it was investigated which had significantly different docking scores (by means of unpaired two-sided Student *t*-tests at significance level of $\alpha = 0.005$). The results are displayed in Table 3.15.

Table 3.15: *At*PARP1 and *At*PARP2 docking score differences

	<i>At</i> PARP1			<i>At</i> PARP2		
	$P < 0.005$	$P > 0.005$	cond. on Δ	$P < 0.005$	$P > 0.005$	cond. on Δ
$\Delta > 0$	1	2	3	20	12	32
$\Delta < 0$	56	18	74	37	20	57
conditional on P	57	20	77	57	32	89

Number of structures docked into *At*PARP1 and *At*PARP2 with significantly different docking scores in comparison to *Hs*PARP1. Results of corresponding Chi-Squared test are shown in 5.6.5

Analysis of docking results

Difference of docking scores of *Hs*PARP1 inhibitors
confidently docked into *At*PARP1 and *At*PARP2

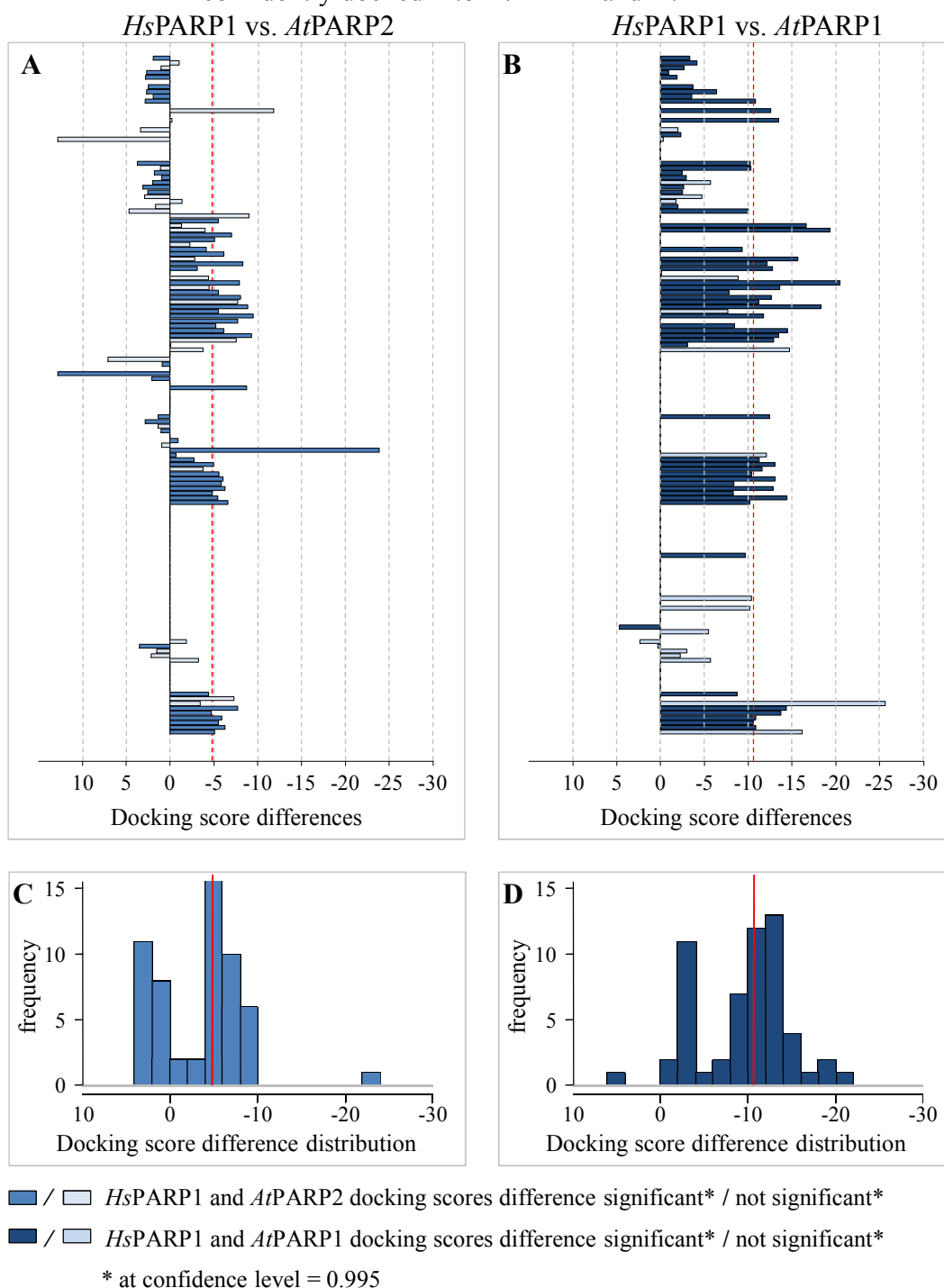


Figure 3.21: *Hs*PARP1 and *Hs*PARP2 docking scores of 142 *Hs*PARP1 inhibitors

A and B: docking analysis of inhibitors docked into *Hs*PARP1. C: docking analysis of inhibitors docked into *Hs*PARP1 and *At*PARP2.

By all of the 77 structures that were docked into the active site of *At*PARP1, only 3 had a positive mean difference of docking scores in comparison to *Hs*PARP1 mean docking scores. Amidst the 74 structures with negative mean differences, 56 of them were significant at

significance level of 0.005. Pearson's Chi-Squared test was used to test for an association of docking scores and statistical significance. The test results are displayed in 5.6.5.

In both cases, there is no evidence against H_0 , therefore the null hypothesis cannot be rejected, indicating that there is an association between the significant score differences and whether these differences are positive or negative. Although the test gives no hint about the type of association, based on the data it can be assumed that if a significant difference between the *Hs*PARP1 and *At*PARP1 (or *At*PARP2) docking score exists, then it is likely that this difference is negative, e.g. the TOTAL SCORE for *Hs*PARP1 is more negative than the TOTAL SCORE for *At*PARP1 or *At*PARP2.

To develop a new docking score threshold for *At*PARP1 and *At*PARP2 based on the mean differences of docking scores and the Pearson's Chi-Squared Test, the medians of the 57 significant score differences between *At*PARP1 and *Hs*PARP2 and *At*PARP2 and *Hs*PARP1 were calculated (red lines in Figure 3.21). The median of the significant docking scores for *At*PARP1 was -10.55 and -4.97 for *At*PARP2. These scores represent the differences of docking scores for known human PARP1 inhibitors that are docked into *At*PARP1/2 active sites. Since the medians of differences are derived from structures that could be docked with confidence into the corresponding active sites and account for significant difference in docking scores, they were used to obtain an adjusted docking threshold for *At*PARP1 and *At*PARP2, based on the threshold for *Hs*PARP1. The new thresholds are calculated as:

Table 3.16: Derivation of new docking thresholds for *At*PARP1 and *At*PARP2

	Threshold <i>Hs</i> PARP1 (3.5.5)		Difference to <i>Hs</i> PARP1		Threshold new
<i>At</i> PARP1	-155.45	-	-10.55	=	-144.91
<i>At</i> PARP2	-155.45	-	-4.97	=	-150.49

The adjusted docking thresholds for *At*PARP1/2 are -144.91 and -150.49, respectively. These docking thresholds can be used to either select compounds from a vendor's database that pass the docking threshold or classify selected compounds as potentially active or inactive based on their docking score.

3.6 Processing data from a commercial data base

The database of Bionet KeyOrganics (2.1.2) was used as commercial database from which potential *At*PARP inhibitors ought to be selected after the virtual screening process.

The database contained 43.179 structures, which were processed with MOE (2.3.1) to be used as input structures for LigPrep.¹⁸⁴ The resulting numbers of protomers and tautomers of the Bionet database after using LigPrep (2.3.5) was 57.117. Of those entries, the three-dimensional structures were generated using ConfGen 2.3.4.²⁴⁷ In total, the number of conformers from 43.179 structures was 1.035.499. These conformations were read into MOE's .mdb database format (2.3.1) and compressed so that it could be used for pharmacophore searches (2.5).

Since the Novikov sample data set contained a high amount of structures that share either a phthalazinone or quinazolinone core, both mimicking the nicotinamid-substructure of NAD⁺, two substructure searches at shop.keyorganics.co.uk for those cores were performed as is displayed in Figure 5.10

These substructure searches resulted in 59 structures sharing a quinazolinone core and 41 structures having a phthalazinone core. This data set of 100 structures was used in addition to the complete database to investigate the docking and pharmacophore performance. Since one of the assumptions was that quinazolinone or phthalazinone core-containing compounds inhibit *Hs*PARP1 and *At*PARP, these structures were used as a reference. It was estimated an overlap between these 100 structures and the large set of the database. The outcomes of the substructure searches were processed with LigPrep (2.3.5) to calculate tautomeric structures. These were also used for pharmacophore filtering and docked into the active site of *At*PARP1.

3.7 Pharmacophore filtering

3.7.1 Pharmacophore selectivity

The pharmacophore was created for the protein models of *Hs*PARP1 and *At*PARP2 (2.5). The selectivity of the pharmacophore was assessed by calculating the percentage of structures that pass the pharmacophore filter. The known *Hs*PARP1 inhibitors from the Novikov data set (2.1.3) and the PARP decoy data set from DUD (2.1.4) were used. The hypothesis was checked whether the pharmacophore could filter out high percentages of decoy structures while retaining the majority of known *Hs*PARP1-inhibiting structures. The results are shown in Table 3.17. Both *Hs*PARP1 and *At*PARP2 pharmacophores filter out about 94% (keeping about 6%) of all decoy structures. Also more than 85% of *Hs*PARP1 inhibitors pass the pharmacophore filter. The percentages of both known inhibitors and decoys that pass the filter are higher for the *At*PARP model than for *Hs*PARP1. The quotient of both fractions can be expressed as the pharmacophore enrichment factor (PEF). A PEF can be understood as the enrichment (here, through a pharmacophore filter) of the fraction of active (non-decoy) structures in a dataset in comparison to the original fraction of actives. The PEF numbers are similar for both models and again slightly higher for the plant model. For a database containing the same number of potential binders and non-binders, the pharmacophore would enrich the fraction of ligands that pass the filter by a factor of about 15 in comparison to the fraction of decoy structures.

Table 3.17: PARP Pharmacophore selectivity

	<i>Hs</i> PARP1 pharmacophore		<i>At</i> PARP2 pharmacophore	
	structures passed (absolute and %)		structures passed (absolute and %)	
<i>Hs</i> PARP1 ligands	121 of 142	85.21	131 of 142	92.25
<i>Hs</i> PARP1 decoys	80 of 1351	5.92	82 of 1351	6.07
Total	201 of 1493	13.46	213 of 1493	14.27
PEF		14.39 *		15.20 **

$$* \text{ PEF}_{\text{pharmacophore } HsPARP1} = \frac{\binom{121}{142}}{\binom{80}{1351}} = 14.39$$

$$** \text{ PEF}_{\text{pharmacophore } HsPARP1} = \frac{\binom{131}{142}}{\binom{82}{1351}} = 15.20$$

3.7.2 Pharmacophore filtering of a commercial database

After having established that the filtering characteristics of the PARP pharmacophore are similar for *At*PARP2 as for *Hs*PARP1 (Table 3.17), the pharmacophore was used to screen a commercial database for potential new *At*PARP inhibitors. The pharmacophore search reduced the structures in the KeyOrganics database from 43.179 to 2.879 tautomeric structures and 2.713 unique structures. This corresponds to a reduction of the data set of about 93.6 %. This value is even higher than the 85.2 % of the *Hs*PARP1 data set for the pharmacophore filter of the *At*PARP2 model (Table 3.17)

3.7.3 Pharmacophore filtering of structures with specific core structures

The 59 and 41 structures having a quinazolinone and phthalazinone core (3.6) were also subjected to *At*PARP2 pharmacophore filtering. By the whole of those 100 structures, 82 passed the pharmacophore (2.5). Structures that did not pass the pharmacophore are substituted in 6- or 7-position and are already known to lower the potency of *Hs*PARP inhibition due to steric clashes in the active site. These results gave further indication of the successful applicability of the pharmacophore.

3.8 Selection of compounds

3.8.1 Compound selection based on docking score and pharmacophore selection

The 2.879 tautomeric (2.713 unique) structures that passed the *At*PARP2 pharmacophore filter, were docked with the PLANTS docking protocol II and ChemPLP scoring function into the active site of *At*PARP2. Due to time restrictions, only 5 instead of 10 solutions per ligand were produced. The 14.395 solutions were checked for the presence of the hydrogen bond whose weight was reduced to the standard value of 1. A hydrogen bond was said to be present if the hydrogen bond was present >30% according to PLANTS scoring function (5.6.8). In association with all docking solutions, those were kept that passed the hydrogen bond filter. Based on the 300 structures with the best (e.g. most negative) docking scores, 136 structures were selected after visual inspection, of which 121 compounds were available at KeyOrganics. Out of those 136 structures were 34 and 31 which contain quinazolinone or phthalazinone cores, while 71 structures contain core structures different from phthalazinones and quinazolinones.

Table 3.18: Compound selection: selected structures and their availability

	QUIN	PHTH	neither QUIN nor PHTH	total
selected	34	31	71	136
available	28	26	67	121

3.8.2 Compound selection based on chemical characteristics

The visual inspection was focussed on general characteristics of structures that were derived from the conformation of FRQ (24, Figure 2.10) bound in the homology model of *At*PARP1 (Figure 3.2). There, the ligand covers the nicotinamide binding site, and a hydrophobic end group is connected to a linker. The ligand does not reach the protein surface or the volume at which the catalytic reaction is assumed to take place. Based on these four characteristics, the ligand binding site was divided into four subsites ($S_1 - S_4$) that correspond to the volumes where the NA binding site (S_1), a linker (S_2) and a hydrophobic region (S_3) could be occupied by a new inhibitor. A fourth region (S_4) represents the protein surface where parts of the acceptor structure would bind. A schematic representation of the active site and two examples of selected structures (25 and 26) are shown in Figure 3.22.

General characteristics of structures selected by VS

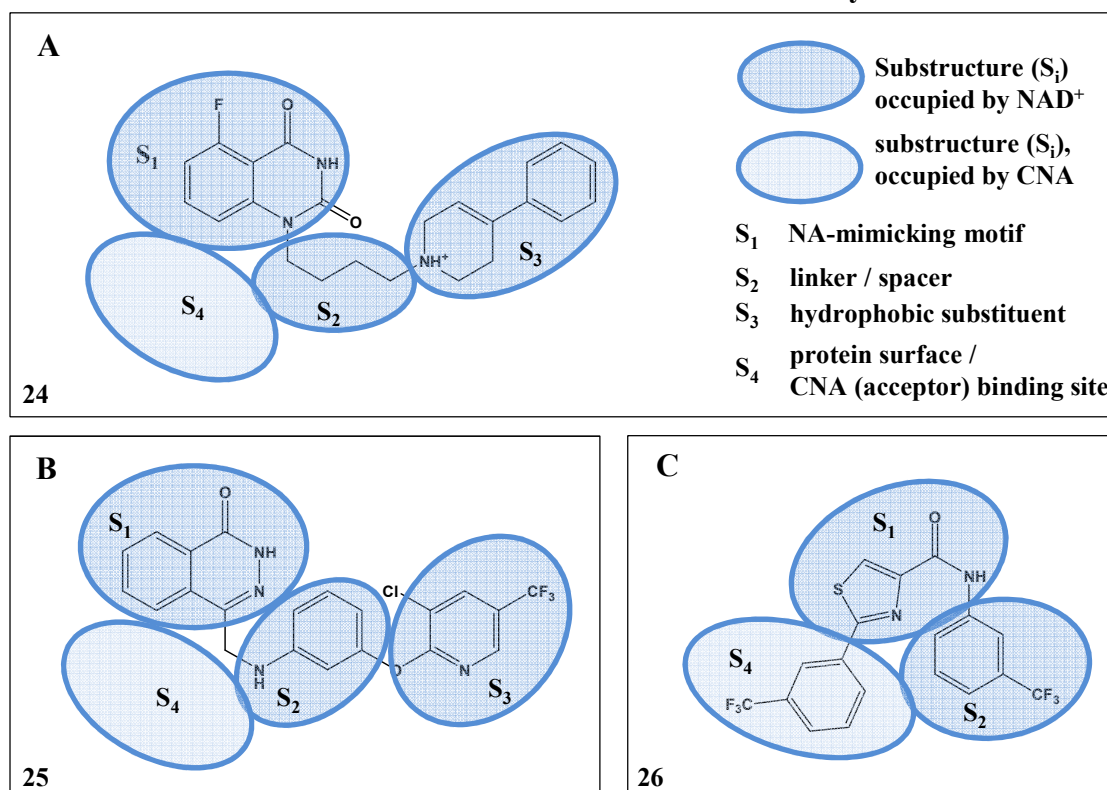


Figure 3.22: Schematic representation of subsites $S1 - S4$ and their occupancies
 subsite representation of A:24, FRQ B:25, Bionet name 10E-62 C:26, Bionet name 12F-408S (see 5.9)

Examples of the structures having a quinazolinone structure are shown in Table 3.19. All those structures belong to quinazolin-4(3H)-one structures that are substituted in 2-position.

Table 3.19: Structures selected for virtual screening: quinazolinones

27	28	29	30	31
32	33	34	35	36
37	38	39	40	41
42	43	44	45	46

Most of these structures carry a hydrophobic substituent that is connected to the quinazolinone core via a methylthio (Table 3.19, structures 27, 32, 37 and 42), ethylene (Table 3.19, structures 28, 33, 38 and 43), isopropyl (Table 3.19, structures 29, 34, 39 and 44), aminoethyl (Table 3.19, structures 30, 35, 40 and 45) or methylene linker (Table 3.19, structures 31, 36, 41 and 46).

The quinazolinone structures were selected such that they form homologous series. So differ quinazolinone compounds 28, 33 and 38 only in the position of the chlorine which is in ortho- meta or para position of the benzyl side chain. Structures 30 and 33 share the same substituent (phenyl ring with a para-substituted chlorine) but differ in the linker. It was also checked for bioisosteric structures, e.g. structures with similar substituents that would influence the physical or chemical properties of the compounds not too much.

Examples of the structures having a phthalazinone structure are shown in Table 3.20. All those structures belong to phthalazin-1(2*H*)-one structures that are substituted in 4-position.

Table 3.20: Structures selected for virtual screening: phthalazinones

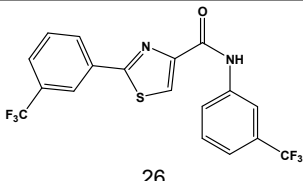
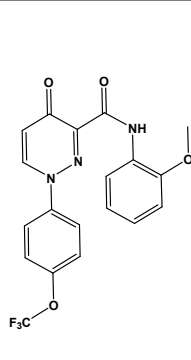
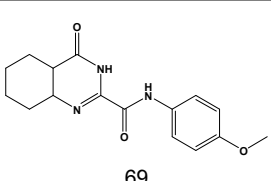
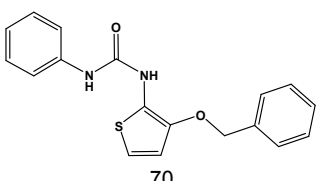
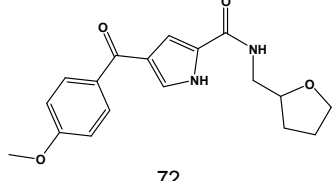
47	48	49	50	51
52	53	54	55	56
57	58	25	59	60
61	62	66	63	68
64	65		67	

These structures carry a hydrophobic substituent that is connected to the phthalazinone core by an aminoethylene, ethylene or methylene linker. Again, structures were selected that share

the same linker but are different in their, mostly hydrophobic, substituent. So are structures 48, 53, 57, 62 and 65 different only in their group present in 4-position of the phenyl ring. Furthermore, 53 and 58 have a trifluoromethyl group attached on the phenyl ring, but in meta- and para-position, respectively.

It was also checked for bioisosteric structures, e.g. structures with similar substituents that would influence the physical or chemical properties of the compounds not too much. Unfortunately, no phthalazinone or quinazolinone structures were found that share the same side chain. This would have been desirable since it would allow to draw conclusions about the influence of the core (phthalazinone or quinazolinone) structure on inhibitor binding.

Table 3.21: Structures selected for *in vitro* screening: other chemical classes

 26	 71	 69
 70		 72

The structures that were selected that lack the phthalazinone or quinazolinone core were divided into classes. Most notably, 11 structures were selected that share an tetrahydroquinazolinone core that is substituted in 2-position by an acetamide moiety. This makes them structurally similar to the class of quinazolin-4(3*H*)-ones. An example of those structures (71) is displayed in Table 3.21. Other classes differ from the substructure scheme as described in A and B. They consist of a substituted 5-membered ring that acts as an analogue of the nicotinamide-mimicking moiety. These structures are substituted in two ring positions; the substituents occupy the linker and hydrophobic regions of the active site (S_2 and S_3), and in some cases target the acceptor site (S_4). Examples of these classes are shown in Figure 3.22: one of those carrying an 1-(3-methoxythiophen-2-yl)urea moiety (72). Structures of these classes are shown in Table 3.21. Another class were structures that share an oxypyridazine core (as 73) that is substituted in 1- and 3-position.

As for the phthalazinone and quinazolinone classes, the structures were selected upon forming homologue series or being bioisosteric to each other.

3.8.3 Analysis of docking results / retrospective power analysis

All 121 available compounds were selected on the basis of *At*PARP2 dockings. Parallel to the *in silico* selection process, Dr. Silky Pienkny was successful in cloning and purifying the catalytic domain of *At*PARP1 (unpublished, data not shown). Furthermore, Dr. Torsten Geissler adapted the *Hs*PARP1 inhibition assay from Trevigen to be used for IC₅₀ calculations of inhibitors of *At*PARP1.^{207,208} Therefore, the structures of the 121 commercially available compounds were docked once again into the active site of *At*PARP1, resulting in different docking scores than in *At*PARP2. In contrast to docking scores of the selected structures, that all would have passed the *At*PARP2 docking threshold, some *At*PARP1-docked structures did not pass the *At*PARP1-specific docking score threshold anymore. Among the *At*PARP1-docking scores of 121 structures, 63 had a docking score < -144.91 and therefore passing the threshold for being classified *in silico* as active, while 58 structures would be classified as inactive according to their docking score > -144.91.

The 121 compounds were tested *in vitro* for their inhibitory activity. As a first *in vitro* test, a compound was classified as active when it reduces *At*PARP1 activity to a level of 60% or less at substrate and inhibitor concentrations of both 100 μM. In classifying compounds as active or inactive, the test allowed a comparison of *in silico* predictions with *in vitro* results which were evaluated by retrospective power analysis.

Table 3.22: Retrospective power analysis

	no. of compounds	<i>in vitro</i> positive	<i>in vitro</i> negative
<i>in silico</i> active (score < -144.91)	63	33	44
<i>in silico</i> inactive (score > -144.91)	58	14	30
total	121	47	74

Among the 121 compounds, 47 proved to be active in *in vitro* tests. By all of those, 33 of them were also predicted to be active, based on the docking score, which corresponds to true positive outcomes. Out of 74 compounds that did not reduce *At*PARP1 activity to more than 40%, 30 were also predicted as inactive based on the docking scores. This fraction represents true negative outcomes. Therefore, the (in)activity of 63 out of 121 compounds was correctly predicted by the docking score, corresponding to a correct decision of outcomes in more than 50%. The number of 44 *in vitro* negative compounds that were incorrectly predicted to be active based in the docking score, correspond to false positive (type II error) outcomes. False negative outcomes are 14 compounds that were shown to inhibit *At*PARP1 more than 40% while having been predicted to be inactive (type I error outcomes).

Retrospective power analysis reveals that the overall docking power is 40.5% (or 30/74). Class-specific power of the docking score-based discrimination ranges from 0% (for the phthalazinones) to 67.2% for classes not having quinazolinone or phthalazinone cores. In all classes statistical power is therefore smaller than the predicted 95%. The overall true positive rate ranges from 50% (for quinazolinones) up to 95% (for phthalazinones). The overall true positive rate of 70.2% is greater than the predicted 44.16% that was based on *Hs*PARP1 dockings.

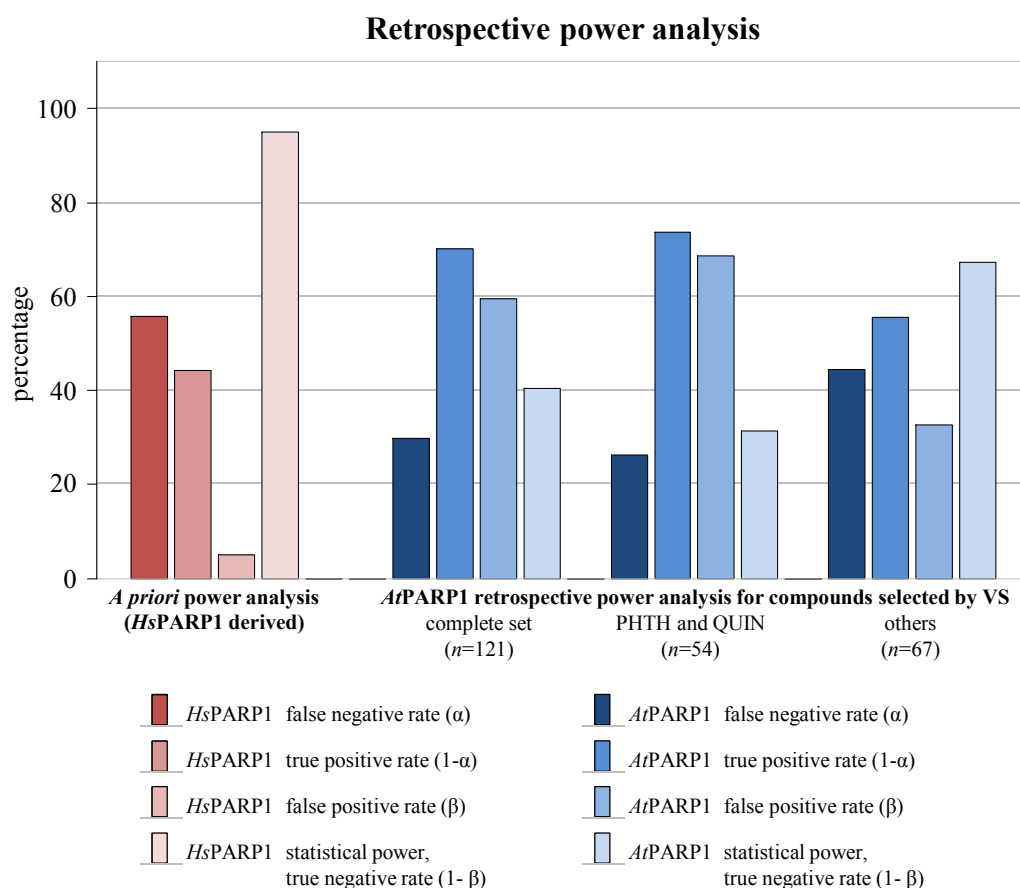


Figure 3.23: Retrospective power analysis
Comparison of observed class-specific error probabilities with predicted error probabilities

A priori power analyses (before *in vitro* data was available) have been run to determine a docking threshold that classifies docked structures as potentially active or inactive against *Hs*PARP1 and subsequently *At*PARP1. The distributions of docking scores for the test set data sets overlap in the region of the calculated docking score threshold (Figure 3.19). These overlaps in the *Hs*PARP1 docking score distributions for the *Hs*PARP1 ligands and decoys data set result in classifying known inhibitor structures as false negative or classifying known decoy structures as false positives with error probabilities of $\alpha = 47\%$ and $\beta = 5\%$. Having transferred this *Hs*PARP1 docking score threshold onto *At*PARP1 virtual screening, several

qualifiers have been imposed that may have influenced the observed *At*PARP1 screening power, therefore it is very likely that predicted *At*PARP1 error probabilities are dissimilar from calculated *Hs*PARP1 error probabilities.

Post hoc, or retrospective, power analyses are discussed controversial. After having observed (mostly insufficient, as in this study) power, researchers do question what might have caused this difference, how much power would have been necessary to detect an effect, or what would have been the minimum sample size for effect detection. This is one of the reasons why some journals even recommend retrospective power analysis.²⁴⁸ In retrospective analyses, one assumes the observed power estimate and true population effect size be the same, but the “observed effect size used to compute the post hoc power estimate might be very different from the true (population) effect size, culminating in a misleading evaluation of power”.²⁴⁹ These differences tend to increase where samples are small and biased.

The paper of Triballeau also points out the character of choosing a too high power or too high docking threshold. The high level of the docking threshold resulting from 95% statistical power “may lend too much credit to the adopted approximations”.²⁴³ From the beginning of the VS strategy to its application on a commercial database, many assumptions and approximations have been set to justify the strategy which was based on *Hs*PARP1 and was transferred to *At*PARP1 and *At*PARP2: those approximations might have led to an observed power of 40.5% for the sample of 121 selected compounds from VS which was much lower than the expected power of 95%. This means that based on 121 selected compounds, only 40.5% of compounds that were shown not to inhibit *At*PARP1 were identified as such, based on the docking score. This discrepancy cannot be explained by retrospective power analysis but there are strong indications for reasons that might have had an influence on decreased power.

The main result of the VS procedure was the identification of compounds that are inhibiting *At*PARP1. In the first *in vitro* screening it was tested if a compound (applied at a 100 μ M concentration) reduces the enzyme activity to less than 40%. From that experiment one could follow if the IC_{40} of an inhibitor would be higher or less than 100 μ M. From that one could estimated that a corresponding IC_{50} would be reached at even higher concentrations. Required enzyme inhibition of 40% was chosen to incorporate inherent errors (biological variability) in the coupled enzyme assay leading to an increased rate of compounds that pass this screening. The concentration of 100 μ M is usually used as the concentration at which a compound is believed to have specific inhibitory effects on the target and hence was set as a threshold.²⁴³

Depending on the target's activity, the size of the database to be screened and the desired endpoint of the VS process, stricter thresholds – e.g. K_i of 5 μM - have been used²⁵⁰ but since the *At*PARP1 in vitro screening of more than 100 compounds was very elaborate and time-consuming, a less strict threshold was chosen: In fact, the choice of a threshold does not only vary from target to target but even for the same target different thresholds can be chosen, depending on the purpose. An example for a wide spectrum for thresholds for one single target is the hERG K⁺ channel where thresholds to discriminate between active and inactive compounds range from 30 μM ²⁵¹ to 300 nM²⁵², covering an activity of three orders of magnitudes.

In this pre-screen, a dichotomous decision concerning the ability of each compound to inhibit *At*PARP could be made and based on this result, a compound was tested more intensively in a second screen in which a compound's IC_{50} value would be determined.

In this study, 47 compounds were identified as active in this test, corresponding to a hit rate of 38.8%. There are only few studies in which VS hit rates in drug discovery are reported. But for those the hit rates are in the range of 20%.^{250,253–255} and this rate is regarded as “respectable”²⁵⁶, especially when the hits were identified for a target whose three-dimensional structure was determined by homology modeling. If the docking threshold is taken into account, 33 compounds of 47 were correctly predicted as active (true positive rate) which would correspond to a docking threshold-specific hit rate of 70.2%.

In the conducted study where 47 out of 121 selected compounds were positively tested for *At*PARP1 inhibitory effects, the costs for testing these compounds were 3630€, a fraction of 38.8% (1410€) resulted in the identification of new *At*PARP1 inhibitors. In comparison to the different screening routes from Table 3.13, this hit rate would correspond to a β -focused screening strategy, because not only the number of compounds was moderate (and therefore the total costs of measuring) but also the fraction of the costs that resulted in positive hits was >38%.

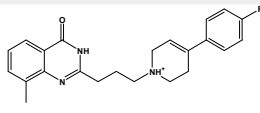
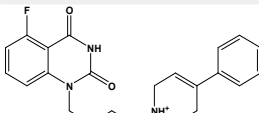
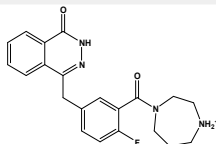
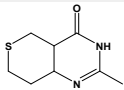
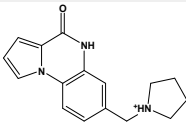
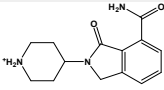
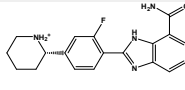
3.9 Finding most probable pose of active compounds

In a first *in vitro* test, each of the 121 available compounds was tested for inhibitory effects on *At*PARP1. Based on these results, 52 compounds were selected for further *in vitro* testing and their IC_{50} values were determined as part of the PhD thesis of Dr. Torsten Geißler.²⁰⁷

To derive (quantitative or binary) structure-activity relationship for the inhibitors based on three-dimensional properties, it is important to know the bioactive conformation of inhibitors. The POSIT workflow was used to find most probable poses of the active compounds inhibiting *At*PARP1. Among the PARP X-ray structures available at PDB that contain co-crystallised inhibitors (April 2012), 18 human and chicken PARP1-3 were selected as templates. To make use of the POSIT workflow, three assumptions were imposed:

- selected PARP X-ray structures contain inhibitors in bioactive conformations
- selected human and chicken PARP1-3 inhibitors also inhibit *At*PARP1 through the same mode of action
- similar conformations (with low RMSD values) for co-crystallised human and chicken PARP1-3 inhibitors would occur for the corresponding inhibitors in *At*PARP1.

Table 3.23: PARP crystal structures used with POSIT

Template (PDB code)	1UK0	1UK1	3C49	
PARP	<i>Hs</i> PARP1	<i>Hs</i> PARP1	<i>Hs</i> PARP3	
resolution (Å)	3.0	3.0	2.8	
inhibitor PDB identifier	FRM	FRQ	KU8	
				
activity data	IC_{50} : 8.0 nM	IC_{50} : 60 nM	IC_{50} : 70 nM	
number of times used	2	1	7	
Template (PDB code)	3C4H	3GJW	3L3L	3L3M
PARP	<i>Hs</i> PARP3	<i>Hs</i> PARP1	<i>Hs</i> PARP1	<i>Hs</i> PARP1
resolution (Å)	2.1	2.3	2.5	2.5
inhibitor contained	DRL	GJW	L3L	A92
				
activity data	K_d : 2000 nM	IC_{50} : 5 nM	EC_{50} : 51 nM	EC_{50} : 6 nM
number of times used	20	20	1	1

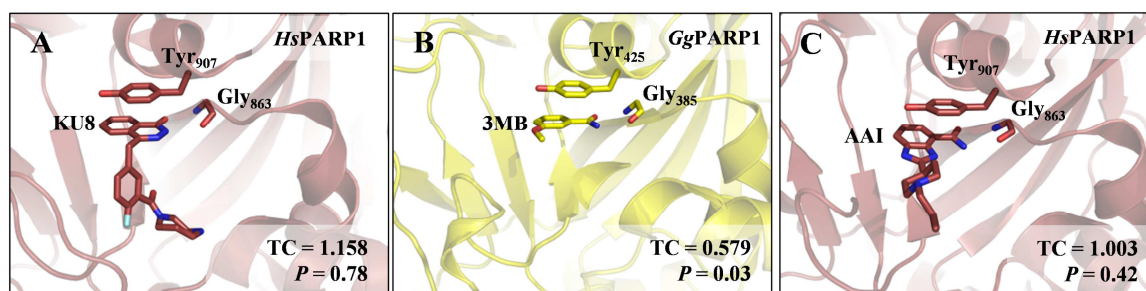
POSIT requires all complexes to search against to have 100% sequence identity and similar

3D-template structures, therefore *At*PARP1 homology models of the 18 PARP template X-ray structures were generated first using the procedure described in 2.3.3.1.

Running POSIT with 18 *At*PARP1 homology models derived from human and chicken PARP1-3 X-ray structures, 52 verified *At*PARP1 inhibitors and the settings as described in 2.3.2, probable bioactive pose(s) for each inhibitor were calculated based on bioactive conformations of human and chicken PARP1-3 inhibitors. In total, 7 out of 18 *Hs*PARP1 and *Hs*PARP3 template structures were used to calculate probabilities for bioactive conformations of new *At*PARP1 inhibitors. These inhibitors are described in Table 3.23. In total, 166 probable bioactive conformations of the known 52 *At*PARP1 inhibitors were calculated because for some inhibitors, more than one bioactive conformation was probable. For those, the conformations with the highest probability were kept for further analysis. To search for a probable bioactive *At*PARP1 inhibitor conformation, all 18 template animal PARP1-3 inhibitor conformations (taken from the PDB) were compared with *Tanimoto_{Combo}* as a measure. The calculated probability had to be greater than 0.05 for a optimisation that leads to a “current best” solution. This procedure is exemplified on the inhibitor 8D-003 (5.9, 73).

Workflow for finding probable bioactive poses for the discovered *At*PARP1 inhibitor 8D-003

Bioactive conformations of co-crystallised *Hs*PARP1 and *Gg*PARP1 inhibitors



From the best docking conformation to probable bioactive conformation

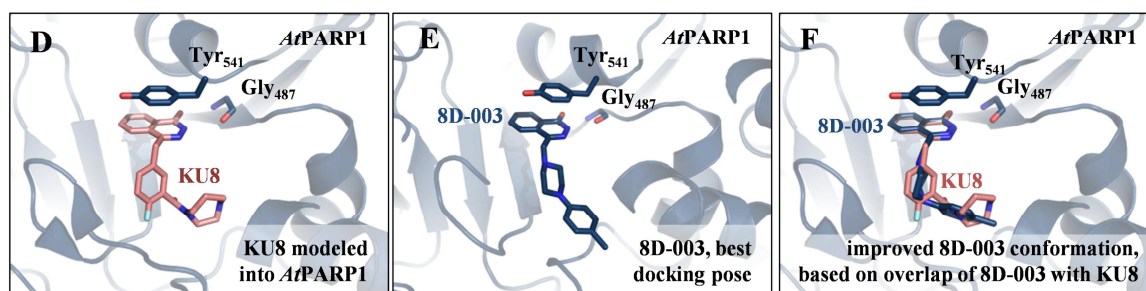


Figure 3.24: Finding probable bioactive conformations of new *At*PARP1 inhibitors
Example with *At*PARP1 inhibitor 8D-003 (5.9, 73) A-C: bioactive conformations of template inhibitors with *Tanimoto_{Combo}* (TC) scores and *P* values. D: template inhibitor with highest TC homology modelled into *At*PARP1, E: the best docking pose of 8D-003 is changed into new pose sharing highest overlap with template inhibitor (F)

POSIT uses generated conformations of the known *At*PARP1 inhibitor to compare against the set of PARP protein-ligand complexes and the *Tanimoto*_{Combo} (TC) score and a probability (P) for each complex (Table 3.24 and Figure 3.24). Out of the 18 template inhibitor conformations, three were rejected because of too low *Tanimoto*_{Combo} scores and probabilities less than 0.05. All other 15 template inhibitors were optimised. The PDB entry 3C49 (representing *Hs*PARP3 with the 4-fluorobenzyl phthalazinone derivative KU8) gave an optimised *Tanimoto*_{Combo} score of 1.310.

Table 3.24: Results of POSIT application I

<i>At</i> PARP1 template	PDB code	Ref:	inhibitor identifier	probability <i>P</i>	<i>Tanimoto</i> combo	Current best
<i>Gg</i> PARP1	1EFY	²⁵⁷	BZC	0.12	0.898	0.912
<i>Hs</i> PARP1	1UK0	⁷⁰	FRM	0.25	0.893	0.893
<i>Hs</i> PARP1	1UK1	¹⁵⁷	FRQ	0.50	1.142	1.201
<i>Gg</i> PARP1	2PAX	⁷²	4AN	0.12	0.739	0.739
<i>Hs</i> PARP1	2RCW		AAI	0.42	1.003	1.058
<i>Hs</i> PARP3	2RD6		78P	0.12	0.866	0.910
<i>Hs</i> PARP3	3C4H	²⁵⁸	DRL	0.42	0.921	0.921
<i>Hs</i> PARP3	3C49	²⁵⁸	KU8	0.78	1.158	1.310
<i>Hs</i> PARP3	3CE0	²⁵⁸	P34	0.42	1.023	1.129
<i>Hs</i> PARP3	3FHB	²⁵⁸	GAB	0.03	0.567	rejected (<i>P</i> <0.05)
<i>Hs</i> PARP1	3GJW	²⁵⁹	GJW	0.25	0.898	1.122
<i>Hs</i> PARP1	3GN7		3GN	0.12	0.875	0.901
<i>Hs</i> PARP2	3KCZ	⁷¹	3AB	0.03	0.579	rejected (<i>P</i> <0.05)
<i>Hs</i> PARP2	3KJD	⁷¹	78P	0.12	0.868	0.884
<i>Hs</i> PARP1	3L3L	²⁶⁰	L3L	0.42	0.904	0.929
<i>Hs</i> PARP1	3L3M	²³⁸	A92	0.50	1.117	1.154
<i>Gg</i> PARP1	3PAX	⁷²	3MB	0.03	0.623	rejected (<i>P</i> <0.05)
<i>Gg</i> PARP1	4PAX	⁷²	NU1	0.09	0.853	0.853

The bioactive conformation of this *Hs*PARP3 inhibitor matches the conformation of the *At*PARP1 inhibitor 8D-003 best. This is also expressed in Figure 3.24.

Using the POSIT workflow, the 52 compounds have a median *Tanimoto*_{Combo} score of 1.29. From 52 structures, 25 (=13+12) structures have a probability *P*>0.5 that the found pose has an RMSD of less than 2 Å to a pose that would be observed in a crystal. For 13 structures this probability is >0.75 meaning that the found pose for those inhibitors is likely to be the correct pose with an RMSD less than 2Å. The median *Tanimoto*_{Combo} score of the subgroup having a quinazolinone structure is higher than the median scores of the subgroup consisting of

phthalazinone substructures or substructures having neither phthalazinone or quinazolinone substructures (Figure 3.25, right panel). Among the 13 structures for which a probability >0.75 was predicted, 12 of them belong to the class of structures sharing a quinazolinone substructure (Figure 3.25).

POSIT results, most probable poses of known *At*PARP1 inhibitors

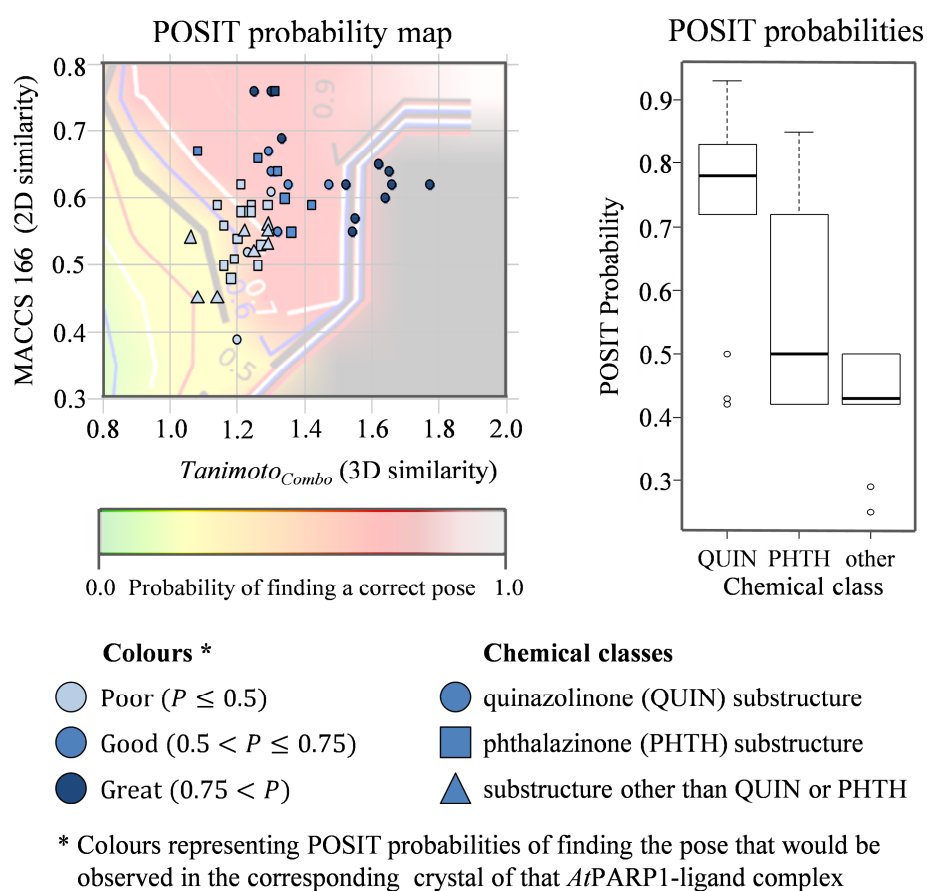


Figure 3.25: POSIT results: quality of prediction for *At*PARP1 inhibitors.

Table 3.25. Results of POSIT application II

characteristic	QUIN	PHTH	other	Total
Number of structures (<i>n</i>)	22	21	9	52
MACCS 166 median	0.62	0.58	0.54	0.58
Tanimoto _{Combo} median	1.34	1.24	1.25	1.29
Probability <i>P</i> median	0.78	0.50	0.43	0.50
number of structures with $P > 0.75$	12	1	0	13
number of structures with $0.50 < P \leq 0.75$	6	6	0	12
number of structures with $P \leq 0.50$	4	14	9	27

From the right panel of Figure 3.25 one can see that the 52 structures concentrate in the upper right part of the POSIT probability map. According to the definition of the POSIT probability

map (2.3.2) the positions of the structures in that map are a result of sufficiently good *Tanimoto_{Combo}* scores of these three-dimensional structures to known crystal structure *Hs/Gg/Mm*PARP inhibitors and an acceptable MACCS166 score that represents two-dimensional feature similarity between the *At*PARP1 inhibitor structures and known *Hs/Gg/Mm*PARP inhibitors. These results give rise to assume that, given the POSIT workflow and the set of PARP crystal structures, the pose of *At*PARP1 inhibitors might be predicted more accurately for quinazolines than for phthalazinones or other structurally similar classes.

The developed VS route focused solely on the identification of potential *At*PARP inhibitors that target the NA part of the active site. This VS route was selected since there is a huge knowledge available about the binding mode of inhibitors that target the NA site. There is only little knowledge about the structural requirements of potential inhibitors that target the AD site of PARP. The reason for that is that there is no crystal structure available that includes NAD⁺, the substrate of PARP, in the active site. While the conformation of NAD⁺ in its bound state has been proposed⁷², the “true” binding of NAD⁺ remains to be elucidated. Protein-bound NAD⁺ conformation in PARP has been inferred from Diphtheria Toxin-like ADP-ribosyltransferases like DT in which NAD⁺ could be crystallised in its bound state.¹²¹ The predicted conformation of NAD⁺ in PARP has not been proven up to 2013, and so are the key interactions responsible for the recognition of the AD site of NAD⁺ remain to be unclear. Because of the (experimental) lack of knowledge about the inhibition of PARP by AD-site inhibitors, the vast majority of PARP inhibitors interact with the NA site.

In 2012, a crystal structure of the human Tankyrase 1 (PDB entry 3UH2)²⁶¹ was released that contains the *Hs*PARP1 inhibitor PJ34.²⁶¹ In this complex, PJ34 is bound in the NA site in a similar conformation as it was observed in the protein-ligand-complexes of *Hs*PARP3 (PDB entry 3C3O)²⁵⁸ and *Hs*PARP15 (PDB entry 3GEY, unpublished) and human Tankyrase 1 (PDB entry 3UH2)²⁶¹. In PDB entry 3UH2, PJ34 is also found in the AD site. The binding mode of PJ34 in the AD site is similar to the binding mode of other Tankyrase inhibitors (e.g. XAV939 in 3KR8).²⁶² More human Tankyrase 1 inhibitors (e.g. IWR2, PDB entry 4DVI)²⁶³ have also been published that bind the AD site. Despite structural differences in the AD binding site of Tankyrases to (human and *Arabidopsis*) PARP1, the release of crystal structures that contain Tankyrase inhibitors that target the AD site, might be a starting point for VS for new classes of PARP inhibitors.

A strategy to screen for PARP inhibitors that target the AD site or mimic the conformation of PARP-bound NAD⁺ might be a solely ligand-based VS approach. For this approach, one

assumes potent AD site-targeting PARP inhibitors to be similar in three-dimensional shape and electrostatic environment to the AD substructure of PARP's substrate NAD^+ . By this reasoning, this approach could be extended to screen databases for NAD^+ -mimicking structures, based on the same assumptions. The three-dimensional shape of NAD^+ is known from the crystal structure of DT-bound NAD^+ ,¹²¹ and its electrostatic environment can be calculated (e.g. using the program EON by OpenEye software). This approach has been successfully applied to identify structures that are similar to the Ca^{2+} -releasing second messenger NAADP.²⁶⁴

Results of VS for structures that are similar in shape and electrostatics to NAADP

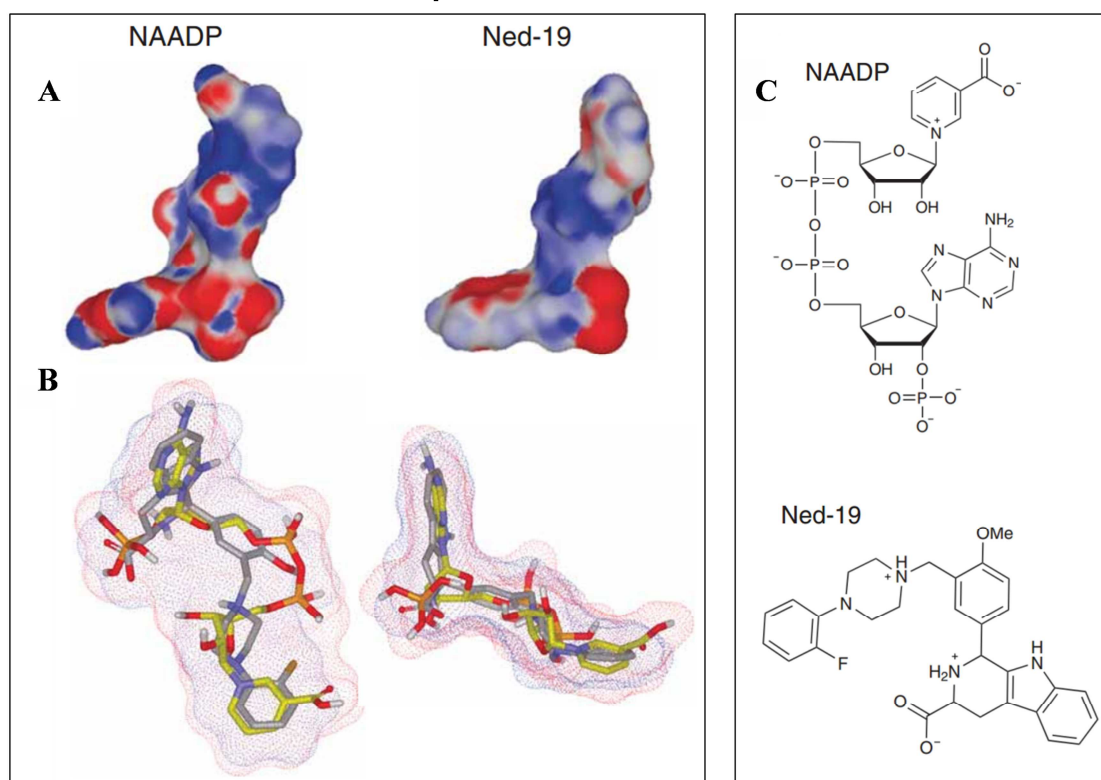


Figure 3.26: POSIT results: quality of prediction for AtPARP1 inhibitors.

A: electrostatic profile of NAADP and Ned-19, B: three-dimensional superposition (shape comparison), of NAADP and Ned-19 C: two-dimensional representation of NAADP and Ned-19

In the study of Naylor and colleagues three-dimensional shapes of NAADP and a database containing 2.7 million structures were calculated with OMEGA. In a second step, the three-dimensional shapes of all database structures were compared to those of NAADP with ROCS and the 500 best hits were saved. Finally the electrostatic overlaps of these hits with NAADP were compared to identify new chemical probes that mimic NAADP. The VS hits were also validated in biological tests. The complete VS route (ranging from two-dimensional representation of 2.7 million structures to calculate three-dimensional conformations,

corresponding electrostatics and comparisons to NAADP) took about 4 months, while the biological testing and validation of less than 50 VS hits took another month. Results of the study are shown in Figure 3.26.

*At*PARP1 inhibitors could also be identified by a modified strategy that combines the applied POSIT workflow (2.3.2) and the natural most probable NAD⁺ conformation in *At*PARP1 as it was described in 2.4.2. This strategy would base on the assumption that the identified conformation of bound NAD⁺ in *At*PARP1 (3.4) would be correct by the definition of POSIT. This assumption could be restated as: The designed model of *At*PARP1 containing NAD⁺ is the same as one would observe in an *At*PARP1-NAD⁺ protein-ligand crystal structure. If this assumption holds, a complete database could be screened for inhibitors that are of similar shape and electrostatics as *At*PARP1-bound NAD⁺. The screening could be focussed on structures that mimic NAD⁺ through similar electrostatics and shape (e.g. having a high Tanimoto Combo score being close to 2 or a POSIT probability > 0.9).

3.10 Structure-activity relationship

3.10.1 General aspects

The conducted docking score analyses for *Hs*PARP1, *At*PARP1 and *At*PARP2 (3.5.6) lead to the identification of new *At*PARP1 inhibitors and the IC₅₀ values for 52 compounds were determined. One aim of the work was to determine which parameters are responsible for the inhibition for *At*PARP1, several approaches have been carried out. In the research of virtual screening for plant PARP inhibitors, two effects were investigated. One was *in vitro* inhibition of *At*PARP1 by competitive inhibitors, the second was the increased dry mass production of *Lolium perenne* L. under induced drought stress conditions which are assumed to be a key factor for increased yield production.

The parameter TOTAL SCORE in the docking program PLANTS proved to be able to discriminate potential *At*PARP inhibitors from structures that were similar in structure and shape. There was no correlation found between TOTAL SCORE and the inhibitor's activity (expressed as the IC₅₀ or the transformed $-\log(\text{IC}_{50}) = \log(1/\text{IC}_{50})$) values by linear regression ($R^2 < 0.001$, data not shown). Also, multiple linear regression (MLR) on all PLANTS's docking output parameters and the IC₅₀ values was performed. For MLR analysis, the PLANTS docking scores were taken from two approaches: The first was by using the original docking scores of the 52 structures whose IC₅₀ later was determined. In the second approach, the approximated most probable bioactive conformation of the 52 structures, being present in complex with *At*PARP1 as a result of the POSIT workflow (2.3.2, 2.4.3 and 3.9), were used as input for rescoring (only using PLANTS's simplex optimisation routine to adapt for better rescoring with ChemPLP). Both approaches resulted in no correlation ($R < 0.001$, data not shown).

Also all 288 two-dimensional molecular descriptors were calculated and used as variables for partial least squares (PLS) and principal component regression (PCR) in MOE (version 2012.10)¹⁶⁹. In both approaches, the numbers of used principal components (PC) were used in a range from 5 to 10. Furthermore, Orthogonal Projections to Latent Structures (O-PLS) was performed with SIMPCA-P (version 10) with a maximum of 7 PC. In all multivariate data analysis approaches, there were no sufficient correlations found between molecular descriptors and *At*PARP1 activity of inhibitors. Since the majority of inhibitors belong to the classes of quinazolinones ($n=21$) and phthalazinones ($n=22$), class-specific MLR, PLS, PCR and O-PLS analyses were performed which also resulted in no correlation (data not shown).

These results were unexpected since QSAR studies with PARP *Hs*PARP1 inhibitors have successfully been applied over the last decade with overwhelming results.^{128,131,132} The study of Rewatkar and colleagues includes phthalazine derivatives and uses MLR technique (and others) to derive QSAR for the class of phthalazinones as PARP inhibitors.¹³²

A reason for the failure of QSAR might be the narrow range of outcome values. For 26 compounds, an IC_{50} between 10 and 80 μM was determined. For another 20 compounds, IC_{50} values from 1 to 10 μM have been observed. Five compounds show an IC_{50} between 0.5 and 1 μM and a single compound has an IC_{50} of 0.08 μM . Apart from the most active compound, 51 compounds cover a range of activity of 2.1 log units. Results that base on such a narrow range can be hard to interpret and are known as a restricted range problem. Together with the small numbers of chemical classes that are represented by these compounds it might be a reason for the results that were obtained for QSAR.

One solution to this problem would have been to extend the determination of IC_{50} values. On the basis of the first screening results (*At*PARP1 inhibition of more than 40% at 100 μM inhibitor concentration) and a few exceptions from that criterion, 52 compounds were selected to IC_{50} value determination. If a compound inhibits *At*PARP1 at 40% at a 100 μM concentration, the IC_{50} value for this compound would be deduced to be in the range between 100 and 500 μM , which might be of no practical relevance anymore.

Also, most compounds identified in the first round of *in vitro* screening would be active in the milli- or micromolar range, whereas the probability of finding a more active – active in the nanomolar range - compound would be very small. For compounds that inhibit *At*PARP1 at 100 μM less than 40%, the corresponding IC_{50} values would be expected to be in the upper micromolar or millimolar range. The determination of IC_{50} values from those compounds would have increased the range of observed IC_{50} values from about 2 log units to 4 or 5 log units which would have eliminated the restricted range problem. It also would have been likely that within the additional set of compounds, different chemical classes of *At*PARP1 inhibitors would have been observed. Both would be two major factors in obtaining better QSAR models. Determining additional IC_{50} values would have been time-consuming and cost-intensive. But this problem could also have been solved by not performing a first screen before verifying individual IC_{50} values but measuring this value without any pre-screens.

3.10.2 Binary structure-activity relationship

Binary quantitative structure-activity relationship (binary QSAR) was used to find physicochemical properties of compounds whose inhibitory effect on *At*PARP1 was assessed by *in vitro* experiments. In the initial experiment, a compound used at a concentration of 100 μ M was defined as inhibiting *At*PARP1 at a level of 100% if it inhibited enzyme activity in the same amount as the negative control compound 4AN. 4AN was verified before as a moderate *At*PARP1 inhibitor as a part of the PhD thesis of Dr. Torsten Geißler.²⁰⁷ An *At*PARP1 inhibition level of 0% was defined as no inhibition of *At*PARP1. As described in 3.8.3, all 121 commercially available KeyOrganics compounds were tested for *At*PARP1 inhibition, and an IC_{50} determination of a candidate was measured in follow-up experiments if the initial inhibitory effect was greater than 40%.

To define training sets of active and inactive structures, compounds that had inhibitory effects in the initial screening that were greater than 60% were treated as active, and compounds that showed inhibitory effects in the initial screening that were less than 20% were treated as inactive. This definition resulted in a selection of 33 active and 58 inactive compounds ($n_{\text{active}}=33$, $n_{\text{inactive}}=58$), respectively. These two data sets were split into training and test sets in a 1:4 ratio. Using this ratio, the training set consisted of 24 active and 44 inactive structures. The test set consisted of 9 active and 14 inactive compounds. (Figure 3.27)

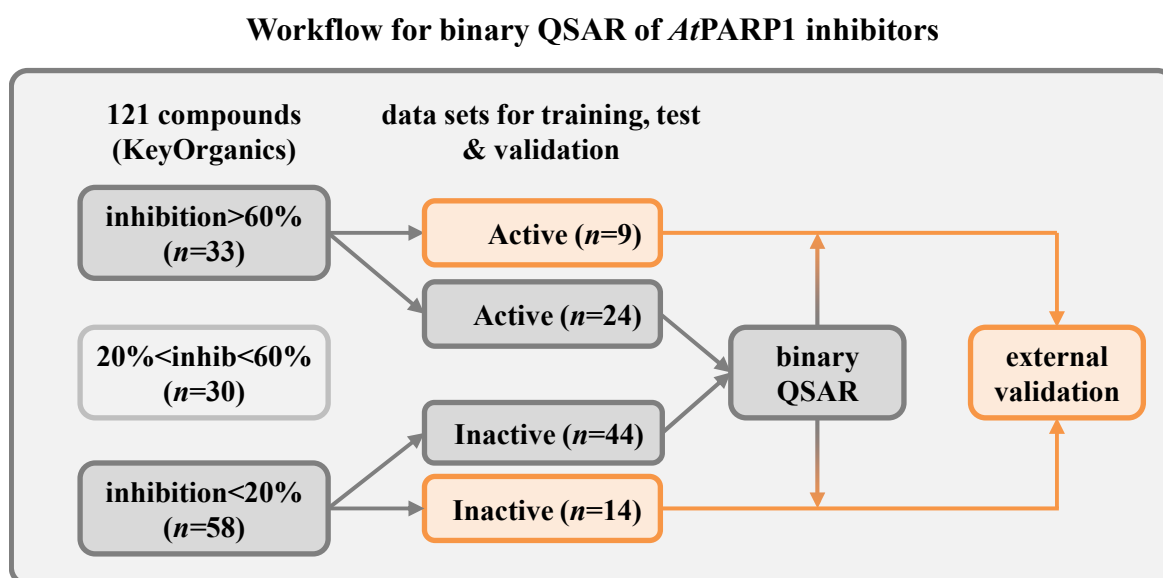


Figure 3.27: Binary QSAR workflow

The number of active and inactive compounds in the test set and validation set in that study are in accordance with the numbers that have been used by others. There are studies where more than 400 structures have been used in the training set²²² and there are binary QSAR

studies conducted with less than 100 structures.²⁵⁶ In the study of Zhou, the number of actives and inactives in the training set was nearly identical ($n_{active}=36$, $n_{inactive}=51$) to the number of structures used in this study.²⁶⁵ The number of compounds for external validation of this study ($n_{active}=9$, $n_{inactive}=14$) is also in accordance with the numbers that have been used before. High numbers of active and inactive structures are desired because larger sample sizes will improve the model quality. As it was used in the study of Zhou and colleagues, the validation set size can also be small ($n_{active}=3$, $n_{inactive}=3$)²⁶⁵ which then poses the question of how well the external test set model accuracy would be for a larger set of validation structures.

Since the bioactive conformations of the training and test set compounds in *At*PARP1 are unknown, two-dimensional descriptors, all available in MOE (version 2012.10)¹⁶⁹, were used for establishing a binary QSAR. By all of the 2D descriptors, the subsets of 16 physical descriptors and 30 partial charge descriptors were selected for model generation. AM1 partial charges were calculated prior to model generation. The following values for model generation were used: binary threshold = 0.5, smoothing parameter = 0.10, condition limit = 10^6 . The number of components was changed from maximum to 5. Among the 46 descriptors, the following 8 were selected for QSAR:

Table 3.26: Selected descriptors for binary QSAR

Descriptor code	Description	rel. importance
Q_VSA_POL	Total polar van der Waals (vdW) surface area	0.1467
Q_VSA_HYD	Total hydrophobic vdW surface area	0.1665
Q_VSA_POS	Total positive vdW surface area	0.2344
Q_VSA_NEG	Total negative vdW surface area	0.2452
Q_VSA_PPOS	Fractional positive vdW surface area	0.1785
Q_VSA_PNEG	Fractional negative vdW surface area	0.1720
Vdw_area	Van der Waals surface area (Å^2) *	0.1546
Vdw_vol	Van der Waals volume (Å^3) *	0.1667

* (calculated using a connection table approximation)

All descriptors calculate specific van der Waals (vdW) surface areas or the vdW volume of a structure. Although they belong to two different classes of 2D descriptors according to MOE classification, they can all be classified into two-dimensional vdW descriptors. Although MOE offers nearly 200 two-dimensional descriptors, there are mainly two ways of selecting relevant descriptors. On the one hand, important descriptors can be selected by performing variable selection methods like PCA.²⁶⁶ On the other hand one can specifically select a subset of descriptor classes that might be important for discriminating actives from inactives.²⁶⁷

The number of 5 principal components (PC) was chosen because the model has a low RMSE (root mean squared error, 5.10.5 and 5.10.8). Two PC showed even lower RMSE but gave worse accuracy in the training set (5.10.5). The smoothing parameter of 0.10 led to slightly worse accuracy than the smoothing factor of 0.05. Increasing the smoothing factor from 0.05 to 0.10 reduces the chance of overfitting while decreasing accuracy only at a low level (5.10.5).

Table 3.27: Results of binary QSAR

	Accuracy in %			LOO cross-validated accuracy in %		
	Total (A)	actives (A ₁)	inactives (A ₀)	total (XA)	actives (XA ₁)	inactives (XA ₀)
accuracy	92.64	79.16	100.00	82.35	54.16	97.72
chance *	56.48	27.94	72.05	58.65	20.58	79.41
P-value	$1.80 \cdot 10^{-9}$	$1.80 \cdot 10^{-10}$		$7.21 \cdot 10^{-5}$	$1.16 \cdot 10^{-5}$	

* theoretical accuracy if there was no association between the model and the sample

By using 5 PC, the model has a total accuracy of 92.64 %, meaning that 63 of 68 training set structures (A) are correctly classified as active or inactive. From the 24 training set actives (A₁), 19 are correctly classified, which corresponds to a sensitivity of 79.16%. The model predicts all inactives in the training set (A₀) correctly which equals 100% specificity. After leave-one-out (LOO) cross-validation, the accuracy on the actives (XA₀) has dropped to 54.16%, meaning that 13 of 24 actives are predicted correctly. The LOO cross-validated accuracy on the inactives (XA₁) is 97.72%, meaning that 43 of 44 inactives are still predicted correctly. (Table 3.27 and Figure 3.28)

Depending on the target, the binary threshold, the selected descriptors and the number of compounds, the total accuracies (based on actives and inactives, A) of binary QSAR models are very high (accuracies > 85%). The accuracies for active (A₁) and inactive (A₀) compounds are in the same range, although there is mostly one class that is much better in accuracy than the other one.²⁶⁸ For the LOO cross-validated training set the accuracies (XA, XA₁, and XA₀) are a bit lower than for the non-cross-validated training set. Another measure of model quality is the theoretical accuracy which gives the accuracy if there was no association between the model and the sample. This value should always be far less than the observed accuracies and in the presented binary QSAR model the differences between the chance and observed accuracies range from ~18% (for XA₀) to more than 50% (for XA₁). Together with the probability (which is far less than 0.05) that the results of this binary QSAR model are due to chance, these facts provide overwhelming evidence that the model accuracy is not achieved by chance.

In the external validation, the model predicts 3 of 9 (33.33%) actives correctly while all 14 inactives are predicted correctly (Figure 3.29). While one usually wants to identify active compounds from QSAR models, higher accuracy on inactives than on actives is not uncommon.^{222,256,265,266,268} This is advantageous because in subsequent *in vitro* screenings (where mostly the active compounds are further examined) compounds that would have no activity (and therefore would be of no interest in lead finding) are excluded through the high binary QSAR accuracy on inactives. Furthermore, the ability of correctly identifying inactive compounds can be directly translated into statistical power (2.7.1), which is desirable for several reasons (3.8.3).

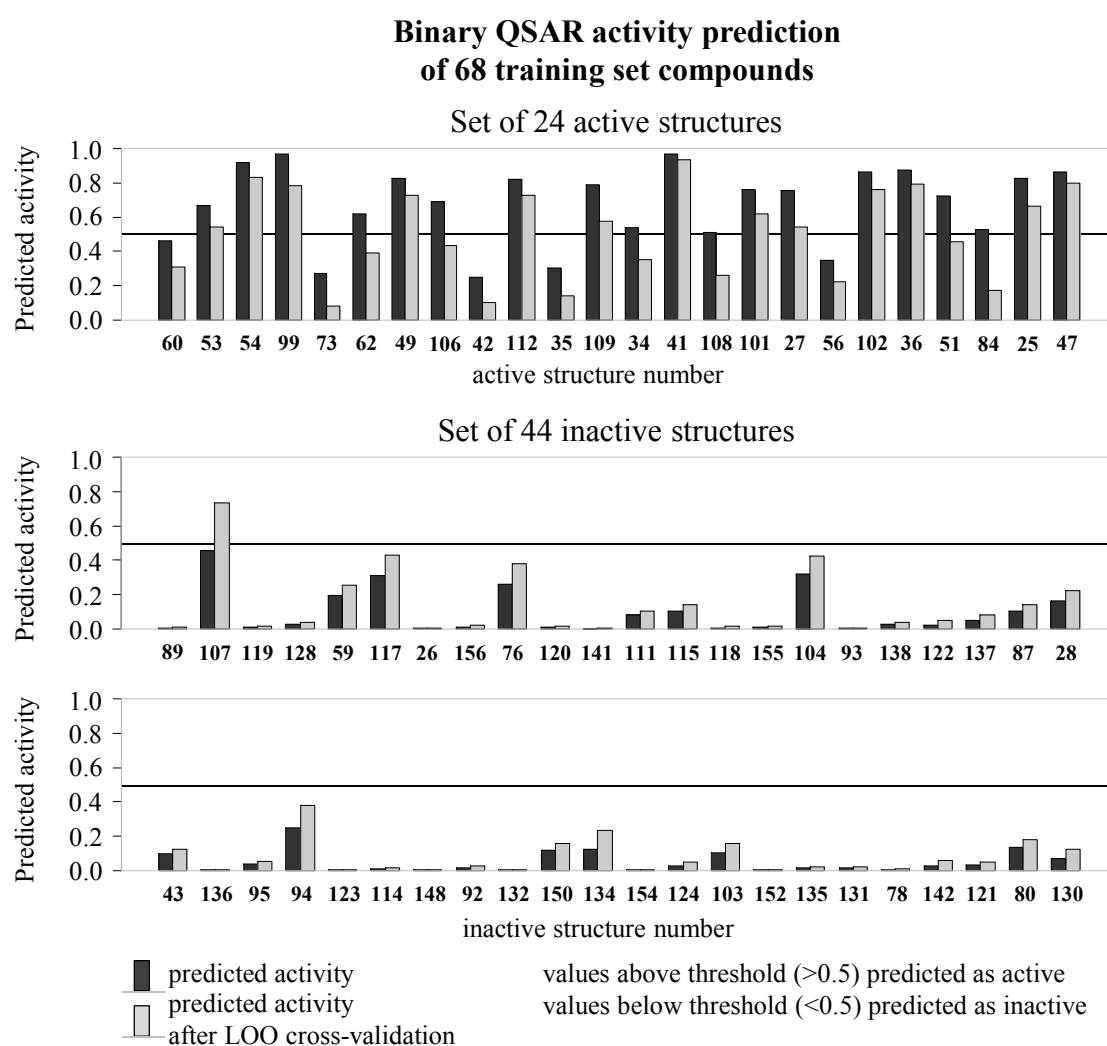


Figure 3.28: Binary QSAR results I

In contrast to QSAR based on continuous outcome values, binary QSAR used binary (e.g. active vs. inactive) outcome values. While obtaining good results from continuous QSAR could probably have been impaired by the restricted range of IC_{50} values and the low amount of chemical class diversity, the restricted range problem is eliminated by assigning a binary

activity value (e.g. 0 or 1) to each of the complete set of compounds. While IC_{50} values were obtained for 52 compounds, a test result in the first screen was available for all 121 selected compounds. Since the activity of a compound in this test was determined in relation to positive and negative controls, each activity value was expected to be in the range of 0% to 100% and the cut-off value was set to 40%. To incorporate measurement and biological variability of the first *in vitro* screen, the binary threshold for selection of the active and inactive training set was defined below 20% below and above 60% *At*PARP1 inhibition at 100 μ M inhibitor concentration. This resulted in the selection of 33 active and 58 inactive compounds. The selection criterion of 20% from the *in vitro* cut-off value of 40% was chosen to avoid QSAR boundary effects. These effects occur for compounds that merely pass or fail to pass the threshold value of 40% inhibition. Because of biological variation, structures having values near the threshold of 40% cannot clearly be assigned to the class of actives or inactives. It has been shown that boundary effect influence the model quality²⁶⁹ and omitting structures with values near the threshold improves the model accuracy²²² and because of that, structures having activity values between 20% and 60% ($n=30$) were excluded from model generation. For the external test set to contain enough compounds, both sets were sorted by activity and each third compound was assigned to the test set.

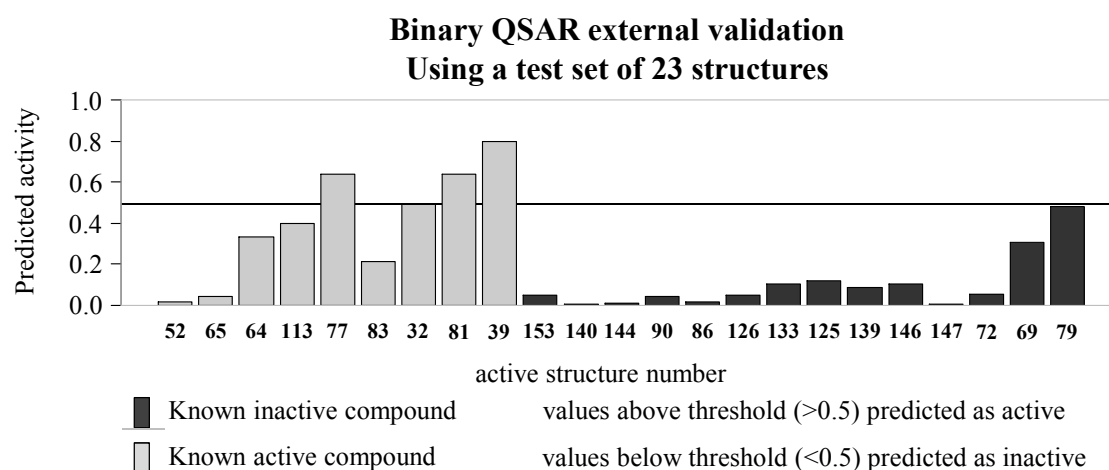


Figure 3.29: Binary QSAR results II – External validation

Binary QSAR was used to find molecular descriptors that can be used to classify *At*PARP1 inhibitors from structures that would not inhibit *At*PARP1. The developed QSAR was specific since it had high accuracy on inactives, both the leave-one-out (LOO) cross-validated and external test set accuracies on inactives were >97%. Since eight two-dimensional descriptors were needed for model generation, this specificity could be used to develop an alternative virtual screening strategy:

To screen a large database (e.g. with more than 1 million unique entries), each being present in 2D coordinates and AM1 partial charges being calculated in advance, the following strategy could be applied: As a first step, the database entries would be filtered with the *At*PARP pharmacophore (2.5). Structures passing the filter will contain already known PARP inhibitor classes like quinazolinone which are likely to be *At*PARP inhibitors. But also structural classes will pass the filter that will not contain PARP inhibitors as has been demonstrated in the first screening test with 121 compounds tested on *At*PARP1 (3.8.3 and 5.10.4). In the conducted VS, 6.28% of the database entries passed the filter. For a million-entry containing database (representing similar fractions of actives and inactives) this percentage would allow more than 60.000 structures to pass this filter. This number is too high to be used for subsequent docking of those structures. But if the selected descriptors hold to discriminate all *At*PARP inhibitors from *At*PARP non-inhibitors for new structures (e.g. the binary QSAR model is as specific for new structures whose *At*PARP activity is unknown) this binary QSAR could be used as a second filter to eliminate structures that are likely not to inhibit *At*PARP.

This strategy was tested on the database in MOE (version 2012.10).¹⁶⁹ That database consists of 653.214 unique structures from commercial vendors. A fraction of 2.84% (18.578 structures) passed the pharmacophore filter. Those structures were then classified by the binary QSAR model. From 18.578 structures, 12.939 structures were classified as potentially inactive (using the threshold of 0.5). Because of the low sensitivity (accuracy on actives) of the binary QSAR model, the remaining 5639 structures cannot be assumed to be *At*PARP1 inhibitors. But the number of remaining structures would now be small enough to be used for molecular docking as it was described in 3.8.3. Another advantage of this strategy is that it would be less time-consuming. Assuming that three-dimensional conformations of structures have already been generated, the pharmacophore search itself on the MOE database was finished within 2 hours. Calculation of AM1 charges as the time-limiting step took a calculation time of a few hours. The prediction of activity using the binary QSAR model was performed in less than a minute. Therefore, a database with one million structures could be screened within one day and the number of structures that pass all filters would be small enough for subsequent docking analysis.

3.10.3 Influence of inhibitors on plant dry mass production

Among the 52 inhibitors for which IC_{50} values were determined, 22 inhibitors (in 26 experiments which included 4 duplicates) were used to study the effects of dry mass production of *Lolium perenne* plants upon drought stress. The inhibitors contained 9 phthalazinone and 7 quinazolinone derivatives, as well as 6 compounds that neither belong to the classes of phthalazinones or quinazolinones. The experiments were conducted by Dr. Heike Hahn at the SKWP Laboratories in Cunnernsdorf. In the experiments, the hypothesis was tested whether stressed plants onto which *At*PARP1 inhibitors were applied show an increased dry mass production in comparison to stressed, but untreated control plants. Drought stress conditions were applied to control plants by reducing the medium's water potential to -0.25 MPa. Drought-stressed plants served as controls ($n=4$) and their dry masses were related to stressed plants. Inhibitors were added to stressed control plants in four different concentrations being 1, 10, 25 and 50 μ M. ($n=4*2=8$)

A general effect caused by the experiment was that the plant's absolute total and shoot dry masses decreased upon being drought-stressed, while the opposite effect was observed for the root dry mass production where the dry mass increased upon drought stress. Additionally, since the dry mass of the shoot accounts for about 90% of the total dry mass, there was a high correlation observed between the dry mass productions of the total dry mass and the shoot dry mass (R^2 of 0.85, 0.85, 0.65 and 0.87 for concentrations of 1, 10, 25 and 50 μ M, respectively). Because of these facts, the shoot dry mass productions was of main interest and analysis. The results are shown in Figure 5.15.

From 22 compounds tested in four different concentrations, 17 showed an increased dry mass production in at least one of the four concentrations. Five compounds showed a positive dry mass growth in three or all four concentrations. Only for one compound (64) a concentration-dependent change in dry mass was observed. Among the 9 phthalazinone derivatives, there were 3 with positive effects in more than two concentrations (48, 49 and 99). From the 7 quinazolinones, one showed positive effects in all four concentrations (46).

For 1 μ M concentration, 4 compounds (49, 129, 46 and 106) showed increased dry mass production, for 10 μ M concentration, 15 compounds showed increased dry mass production, among those were 6 phthalazinone derivatives and 4 quinazolinone derivatives. For 25 μ M concentration, 7 compounds (99, 77, 49, 54, 48, 129 and 46) showed increased dry mass production. For the highest concentration of 50 μ M, 7 compounds (99, 81, 49, 117, 48, 129

and 46) showed increased dry mass production; among those were 6 phthalazinone derivatives and 4 quinazolinone derivatives.

There was no general dose dependency detectable, besides for 64, which showed decreased dry masses for higher concentrations. This effect could not be generalized, e.g. the correlation coefficients between 1 and 10 μM , between 10 and 25 μM and between 25 and 50 μM were 0.11, 0.43, and 0.30, respectively. That means that one cannot derive the dry mass production from one concentration to the next higher concentration. There was no correlation between the IC_{50} values of the inhibitors and the shoot dry mass growths found (with $R^2=0.04$, 0.08, 0.18 and 0.17 for the four concentrations of 1, 10, 25 and 50 μM , data not shown). This indicates that PARP is not the only player in the network of drought stress response.

Currently there is only one whole-plant assay published that measures effects of compounds on drought stress and could be used for high or medium throughput screening. This *Lemna minor* plants assay was developed by our group in which the growth rate of a treated plant is compared to the untreated plant's growth rate upon drought stress application via PEG.²⁰⁷ All 52 compounds for which IC_{50} values for *At*PARP1 inhibition have been determined were subjected to this assay and the observed plant growth of those compounds was compared to their IC_{50} for *At*PARP1. Using a simple linear regression model, there was no correlation found between *Lemna minor* growth and the IC_{50} values ($R^2 < 0.001$, data not shown). This result could be explained by the assumption that *At*PARP1 inhibitors are no inhibitors of *Lemna minor* PARP (which would be equivalent to the term *Lemna minor* decoys) or these compounds inhibit *At*PARP1 and *Lemna minor* PARP through different modes of action. It might also be that PARP is not a relevant target, at all or that secondary effects as compound metabolism interfere with effects that have been observed in *At*PARP *in vitro* / *in planta* studies. If one assumes that those compounds inhibit both enzymes through the same mode of action, there is no evidence that all compounds reach their target at the concentration at which they were applied to the medium in the *Lemna minor* assay. *Lemna* growth might also very likely be influenced by acting of *At*PARP1 inhibitors on other targets, especially on those with similar active sites to PARP or on those that use NAD^+ as substrate or co-substrate.

4 Summary and outlook

In this work, an *in silico* characterisation of the *Arabidopsis thaliana* Poly-(ADP-ribose)-Polymerase (*At*PARP1) and the first virtual screening study for a plant PARP enzyme was conducted, which resulted in the identification of 52 *At*PARP1 inhibitors.

Using a broad range of molecular modelling tools, the catalytic domain of *At*PARP1 was characterised *in silico*. This characterisation encompassed the investigation of protein stability from which it was concluded that the three-dimensional shape of this conserved PARP domain is of high similarity to the *Hs*PARP catalytic domain. Furthermore, there is overwhelming evidence provided by MD simulation of the *At*PARP1 catalytic domain in complex with the natural substrates (or substrate analogues) NAD⁺ and CNA (an ADP-ribose analogue) that *At*PARP1 binds its substrates in an analogous manner as it is described for *Hs*PARP and as it is observed for ADPR-transferases like Diphtheria Toxin. Based on the results of that work, the role of a conserved glutamate essential for the catalytic reaction in PARP is the same in *At*PARP1 as was shown for other ADPRT.

As for the *in silico* characterisation of the catalytic domain of *At*PARP1, the virtual screening for *At*PARP1 inhibitors involved the use of statistical tools like receiver operator characteristics (ROC) curves and power analysis to guide the VS process and improve its quality.

Based on the identification of *At*PARP1 inhibitors but also from proposed compounds which proved not to be *At*PARP1-active, general characteristics (descriptors) of the structures were used to derive binary quantitative structure-activity relationship (QSAR) that could help to understand the structural requirements that are responsible for *At*PARP1 inhibition.

This work contributes to an understanding of the role of *At*PARP1. Since PARP are implicated as a first responder to drought stress (by depleting the NAD⁺ pools of the plant that leads to disturbances in energy homeostasis upon drought stress), inhibitors of these enzymes might increase the drought stress tolerance of plants. To further test this hypothesis, application of identified inhibitors on crop plants like *Zea mais* would be desirable. During this study, some of the inhibitors showed increased dry mass production in *Lolium perenne* giving first hints that these inhibitors do increase the drought stress tolerance in plants; although in this study the effects could not solely be related to PARP enzymes. Further investigations also could involve the ability to selectively inhibit plant PARP enzymes. In

Arabidopsis, it is still not completely investigated if selective PARP inhibitors increase the drought stress tolerance more than unselective ones.

In the same context the selectivity of plant PARP inhibitors with respect to human PARP is also of importance because human PARP inhibitors are promising compounds to treat severe conditions like breast and ovarian cancer or ischemia-reperfusion injury. Since a lot of molecular modelling tools have already been used to study human PARP, computer-aided drug design in the context of the development of selective plant PARP inhibitors would be a useful tool to elucidate the role of PARP and drought stress. Furthermore, from the identified inhibitors, lead compounds could be derived with higher potency or ADME(T) characteristics, using molecular modelling tools in close collaboration with medicinal chemistry.

5.1.11 *At*PARP2 – *Hs*PARP3

1: cd_AtPARP2_633-979; 2: cd_HsPARP3_182-533; Length: 370
 # Identity: 125/370 (33.8%); Similarity: 183/370 (49.5%); Gaps: 41/370 (11.1%); Score: 458.5

cd_AtPARP2_	1	SSNLAPSLIELMKMLFDVETYSAMMEFEINMSEMPGKLSKHNIQKGF	50
cd_HsPARP3_	1	PCSLDPATQKLTITNIFSKEMFKNTMALMDLDVKKMPLGKLSKQIARGFE	50
cd_AtPARP2_	51	ALTEIQRLLTESDPQPTMKESLLVDASNRFFTMIP----SIHPHIIRDED	96
cd_HsPARP3_	51	ALEALEEAL----KGPTDGGQSLEELSSHFYTVIPHNFGHSQPPINSPE	96
cd_AtPARP2_	97	DFKSKVKMLEALQDIEIASRIVGFDVDSTES-----LDDKYKKLHCD	138
cd_HsPARP3_	97	LLQAKDMLLVLADIELAQALQA--VSEQEKTVEEVPHPPLDRDYQLLKQC	144
cd_AtPARP2_	139	ISPLPHDSEDYRLIEKYLNTT----HAPHTTEWSLELEEVFALEREGEFD	184
cd_HsPARP3_	145	LQLLDSGAPEYKVIQTYLEQTGNSHRCPT-----LQHIWKVNQEGEED	187
cd_AtPARP2_	185	KYAPHREKLGKMLLWHGSRLTNFVILNQGRLRIAPPEAPATGYMFGKGI	234
cd_HsPARP3_	188	RFQAH-SKLGKRLWHGHTNMAVVAAILTSLGLRI----MPHSGGRVKGKI	232
cd_AtPARP2_	235	YFADLVSKSAQY--YTCKKNPVGLMLLSEVALGEIHEL-TKAKYMDKPP	281
cd_HsPARP3_	233	YFASENSKSAQYVIGMKCGAHHVGYMFLGEVALGREHHINTDNPSLKSPP	282
cd_AtPARP2_	282	RGKHSKGLGKVP---QDSEFAKWRGDVTVPCGKPV--SSKVKASELMYN	327
cd_HsPARP3_	283	PGFDSVIARGHTEPDTQDTELELDGQQVVVPGQPVPCPEFSSSTFSQS	332
cd_AtPARP2_	328	EYIVYDTAQVKLQFLLKVRF	347
cd_HsPARP3_	333	EYLIYQESQCRRLRYLLEVHL	352

5.1.12 *At*PARP2 – *Gg*PARP1

1: cd_AtPARP2_633-979; 2: cd_GgPARP1_659-1004; Length: 355
 # Identity: 172/355 (48.5%); Similarity: 245/355 (69.0%); Gaps: 17/355 (4.8%); Score: 867.5

cd_AtPARP2_	1	SSNLAPSLIELMKMLFDVETYSAMMEFEINMSEMPGKLSKHNIQKGF	50
cd_GgPARP1_	1	KSKLAKPIQDLIKMIFDVESMKKAMVEFEIDLQKMPGKLSKRQIQSAYS	50
cd_AtPARP2_	51	ALTEIQRLLTESDPQPTMKESLLVDASNRFFTMIPSIH-----PHIIRD	94
cd_GgPARP1_	51	ILNEVQQAQVSDGG-----SESQILDLSNRFYTLIP--HDFGMKKPPLLSN	93
cd_AtPARP2_	95	EDDFKSKVKMLEALQDIEIASRIV--GFDVDSTESLDDKYKKLHCDISPL	142
cd_GgPARP1_	94	LEYIQAKVQMLDNLLEDIEVAYSLLRGGNEDGDKDPIDINYEKLRDIDKVV	143
cd_AtPARP2_	143	PHDSEDYRLIEKYLNTTHAPHTTEWSLELEEVFALEREGEFDKYAPHREK	192
cd_GgPARP1_	144	DKDSEEAIIKQYVKNTHAATHNAYDLKVVEIFRIEREGESQRYKPFKQ-	192
cd_AtPARP2_	193	LGKMLLWHGSRLTNFVILNQGRLRIAPPEAPATGYMFGKGIYFADLVSK	242
cd_GgPARP1_	193	LHNRQLLWHGSRTTNFAGILSQGLRIAPPEAPVTGYMFGKGIYFADMVSK	242
cd_AtPARP2_	243	SAQYCYTCKKNPVGLMLLSEVALGEIHELTKAKYMDKPPRGKHSKGLGK	292
cd_GgPARP1_	243	SANYCHTSQADPIGLILLGEVALGNMYELKNASHITKLPKGKHSVKGLGK	292
cd_AtPARP2_	293	KVPQDSEFAKWRGDVTVPCGKPVSSKVKASELMYNEYIVYDTAQVKLQFL	342
cd_GgPARP1_	293	TAPDPTATTTLDG-VEVPLNGISTGINDTCLLYNEYIVYDVAQVNLKYL	341
cd_AtPARP2_	343	LKVRF	347
cd_GgPARP1_	342	LKLF	346

5.1.15 *At*PARP3 – *Hs*PARP3

1: cd_AtPARP3_449-801; 2: cd_HsPARP3_182-533; Length: 377
 # Identity: 88/377 (23.3%); Similarity: 153/377 (40.6%); Gaps: 49/377 (13.0%); Score: 215.0

cd_AtPARP3_	1	HCKLDSFVANFIKVLGCGQEIYNYALMELGLDPPDLPMGMLTDIHLKRCEE	50
cd_HsPARP3_	1	PCSLDPATQKLTITNIFSKEMFKNTMALMDLDVKKMPLGKLSKQIARGFE	50
cd_AtPARP3_	51	VLLEFVEKVKTTKETGQKAEAMWADFSSRWFSLM-----HSTRPMLRHDV	95
cd_HsPARP3_	51	ALEALEEALKGPTDGGQSLE----ELSSHFYTVIPHNFHGSQPP----PI	92
cd_AtPARP3_	96	N--ELADHAASAFETVRDINTASRLIG----DMRGDTLDDPLSDRYKKLG	139
cd_HsPARP3_	93	NSPELQAKKDMLLVLADIELAQALQAVSEQEKTVEEVPHPLDRDYQLLK	142
cd_AtPARP3_	140	CKISVVDKESEDYKMVVKYLETTYEPVKVSDVEYGVSVQNVFAV----ES	185
cd_HsPARP3_	143	CQLQLDSDGAPEYKVIQTYLEQTGSNHRCP-----TLQHIWKVNVQEGEE	186
cd_AtPARP3_	186	DAIPSLDDIKKLPNKVLLWCGSRSSNLLRHIYKGFPAVCSLPVPGYMF	235
cd_HsPARP3_	187	DRFQA---HSKLGNRKLWHGTNMAVVAAILTSG---LRIMPHSGGRVG	229
cd_AtPARP3_	236	RAIVCSDAAAEAAARY--GFTAVDRPEGFLVLAVASLGEE-----VTEFT	277
cd_HsPARP3_	230	KGIYFASENSKSAGYVIGMKCGAHHVGYMFLGEVALGREHHINTDNPSLK	279
cd_AtPARP3_	278	SPPEDTKTLEDKKGIVKGLGRKKTEESEHFMRDDIKVPCGRLVP--SEHK	326
cd_HsPARP3_	280	SPPPGF---DSVIARGHTEPDPTQDTELELDGQQVVPVQGPVPCPEFS	325
cd_AtPARP3_	327	DSPLEYNEYAVYDPKQTSIRFLVEVKY	353
cd_HsPARP3_	326	SSTFSQSEYLIYQESQCLRLYLLEVHL	352

5.1.16 *At*PARP3 – *Gg*PARP1

1: cd_AtPARP3_449-801; 2: cd_GgPARP1_659-1004; Length: 360
 # Identity: 105/360 (29.2%); Similarity: 182/360 (50.6%); Gaps: 21/360 (5.8%); Score: 361.5

cd_AtPARP3_	1	HCKLDSFVANFIKVLGCGQEIYNYALMELGLDPPDLPMGMLTDIHLKRCEE	50
cd_GgPARP1_	1	KSKLAKPIQDLIKMIFDVESMKKAMVEFEIDLQKMPGKLSKRQIQSAYS	50
cd_AtPARP3_	51	VLLEFVEKVKTTKETGQKAEAMWADFSSRWFSLM-HS---TRPMLRHDVN	96
cd_GgPARP1_	51	ILNEVQQAV---SDGGSSEQIL--DLSNRFYTLIPHDFGMKKPPLLSNL-	94
cd_AtPARP3_	97	ELADHAASAFETVRDINTASRLI--GDMRGDTLDDPLSDRYKKLGCKISV	144
cd_GgPARP1_	95	EYIQAKVQMLDNLDDIEVAYSLLRGGNEDGD--KDPIDINYEKLRTDIKV	142
cd_AtPARP3_	145	VDKESSEDYKMVVKYLETTYEPVKVSDVEYGVSVQNVFAVESDA-IPSLDD	193
cd_GgPARP1_	143	VDKDSEEAIIKQYVKNTH---AATHNAYDLKVVEIFRIEREGESQRYKP	189
cd_AtPARP3_	194	IKKLPNKVLLWCGSRSSNLLRHIYKGFPAVCSLPVPGYMFGRRAIVCSDA	243
cd_GgPARP1_	190	FKQLHNRQLLWHGSRRTTNFAGILSQGLRIAPPEAPVTGYMFGKGIYFADM	239
cd_AtPARP3_	244	AAEAARYGFTAVDRPEGFLVLAVASLGEVTEFTSPPEDTKTLEDKKGIV	293
cd_GgPARP1_	240	VSKSANYCHTSQADPIGLIILGEVALG-NMYELKNASHITK-LPKGKHSV	287
cd_AtPARP3_	294	KGLGRKKTEESEHFMRDDIKVPCGRLVPSEHKDSPLEYNEYAVYDPKQT	343
cd_GgPARP1_	288	KGLG-KTAPDPTATTTLDGVEVPLGNGISTGINDTCLLYNEYIVYDVAQV	336
cd_AtPARP3_	344	SIRFLVEVKY	353
cd_GgPARP1_	337	NLKYLLKLF	346

5.2 Multiple Sequence Alignment

5.2.1 Clustal Omega settings

Input Parameters

```

Program: clustalo; version: 1.2.0;
Output guide tree: false; Output distance matrix: false; Dealign input sequences: false
mBed-like clustering guide tree: true; mBed-like clustering iteration: true
Number of iterations: 0; Maximum guide tree iterations: -1; Maximum HMM iterations: -1;
Output alignment format: clustal; Output order: aligned; Sequence Type protein;

```

5.2.2 Multiple Sequence Alignment I

```

cd_AtPARP2_633-979      SSNLAPSLIELMKMLFDVETYSAMMEFEINMSEMPLGKLSKHNIQKGFALTEIQRLLT
cd_HsPARP1_662-1007    KSKLPKPVQDLIKMIFDVESMKKAMVEYEIDLQKMPGKLSKRQIQAAYSILSEVQQAVS
cd_GgPARP1_659-1004    KSKLAKPIQDLIKMIFDVESMKKAMVEFEIDLQKMPGKLSKRQIQSAYSILNEVQQAVS
cd_AtPARP1_286-633    QSKLDRVAKFISLICNVSMMAQHMEIGYNANKLPLGKISKSTISKGYEVLKRISEVID
cd_HsPARP2_231-577    ESQDLDRVQELIKLICNVQAMEEMMEMKYNTKKAPLGKLTVAQIKAGYQSLKKIEDCIR
.*:*      : ..... :*.      .*:*      :. : *****: * . . . * . . . :

cd_AtPARP2_633-979      ESDPQPTMKESLLVDASNRFFFTMIPSI---HPH--IIRDEDDFKSKVKMLEALQDIEIA
cd_HsPARP1_662-1007    QGS-----SDSQILDLSNRFYTLIPHDFGMKKPP--LLNNADSVQAKVEMLDNLLDIEVA
cd_GgPARP1_659-1004    DGG-----SESQILDLSNRFYTLIPHDFGMKKPP--LLSNLEYIQAKVQLDNLLDIEVA
cd_AtPARP1_286-633    RYD-----R-TRLEELSGEFYTVIPHDFGFKMSQFVIDTPQKQKQIEMVEALGEIELA
cd_HsPARP2_231-577    AGQ-----HGRALMEACNEFYTRIPHDFGLRTPP--LIRTQKELSEKIQLLEALGDIEIA
      : : . .*:* **      : : . . * : : : : * : * : * : *

cd_AtPARP2_633-979      SRIV--GFDVDSTESLDDKYKKLHCDISPLPHDSEDYRLIEKYLNTHAPTHTWESLELE
cd_HsPARP1_662-1007    YSLLRGGSDDSSKDFIDVNYEKLKTDIKVVDRDSEAEIIRKYVKNTHATTHNAYDLEVI
cd_GgPARP1_659-1004    YSLLRGGNEDGDKDFIDINYEKLRTDIKVVDKDSEEAQIKQYVKNTHAATHNAYDLKVV
cd_AtPARP1_286-633    TKLLSVDP--GLQDDPLYHYHQQLNCGLTPVGNDSSEFSMVANYMENTHAKTHSGYVTEIA
cd_HsPARP2_231-577    IKLVKTEL--QSPEHPLDQHYRNLHLCALRPLDHESYEFKVISQYLQSTHAPTHSDYTMTLL
      : :      . : .*:* . : : .* : : : * : : * : * : * : * : *

cd_AtPARP2_633-979      EVFALEREGEFDKYAPHREKLGKMLLWHGSRLTNFVGIHQGLRIAPPEAPATGYMFGK
cd_HsPARP1_662-1007    DIFKIEREGECQRYKPFK--QLHNRLLLWHGSRTTNFAGILSQGLRIAPPEAPVTGYMFGK
cd_GgPARP1_659-1004    EIFRIEREGESQRYKPFK--QLHNRQLLWHGSRTTNFAGILSQGLRIAPPEAPVTGYMFGK
cd_AtPARP1_286-633    QLFRASRAVEADRFQQFS--SSKNRMLLWHGSRLTNWAGILSQGLRIAPPEAPVTGYMFGK
cd_HsPARP2_231-577    DLFEVEKDGEKEAFRE---DLHNRMLLWHGSRMSNVWVILSHGLRIAPPEAPITGYMFGK
:* . : * : : . * : * : * : * : * : * : * : * : * : * : * : *

cd_AtPARP2_633-979      GIYFADLVSKSAQYCYTCKKNPVGLMLLSEVALGEIHELTKAKY-MDKPPRGKHSKTKGLG
cd_HsPARP1_662-1007    GIYFADMVSKSANYCHTSQGDPIGLILLGEVALGNMYELKHASH-ISKLPKGKHSVKGLG
cd_GgPARP1_659-1004    GIYFADMVSKSANYCHTSQADPIGLILLGEVALGNMYELKNASH-ITKLPKGKHSVKGLG
cd_AtPARP1_286-633    GVYFADMFSKSANYCYANTGANDGVLLLCEVALGDMNELLYSDYNADNLPPGKLSKTKGVG
cd_HsPARP2_231-577    GIYFADMSSKSANYCFASRLKNTGLLLLSEVALGQCNELEANPKAEGLLQGHSTKTKGLG
* : * : * : * : * : * : * : * : * : * : * : * : * : * : * : *

cd_AtPARP2_633-979      KKVPQDSEFAKWRGDVTVPCGKPVSSKV---KASELMYNEYIVYDQAQVKLQFLLKVRF
cd_HsPARP1_662-1007    KTTDPDPSANISL-DGVDVPLGTGISSGV---NDTSLLYNEYIVYDIAQVNLKYLLKLF
cd_GgPARP1_659-1004    KTAPDPTATTTL-DGVEVPLNGISTGI---NDTCLLYNEYIVYDVAQVNLKYLLKLF
cd_AtPARP1_286-633    KTAPNPSEAQTLEDGVVPLGKPVVERSC---SKGMLLYNEYIVYVVEQIKMRYVIQVKF
cd_HsPARP2_231-577    KMAPSSAHFVTLN-GSTVPLGPASDTGILNPDGYTLNYNEYIVYVNPQVRMRYLLKVQF
* . * . : . * * * . . * * : * : * : * : * : * : * : * : *

#
# Percent Identity Matrix
#

1: cd_AtPARP2_633-979      100.00      51.18      50.00      42.18      45.10
2: cd_HsPARP1_662-1007    51.18      100.00      86.99      46.22      46.49
3: cd_GgPARP1_659-1004    50.00      86.99      100.00      45.93      47.37
4: cd_AtPARP1_286-633    42.18      46.22      45.93      100.00      50.73
5: cd_HsPARP2_231-577    45.10      46.49      47.37      50.73      100.00

```

5.2.3 Multiple Sequence Alignment II

```

cd_AtPARP3_449-801      HCKLDSFVANFIKVLGCGQEIYNYALMELGLDPPDLPMGMLTDIHLKRCEEVLLFVEKVK
cd_HsPARP3_182-533     PCSLDPATQKLIITNIFSKEMFKNTMALMDLDVKKMPLGKLSKQQIARGFEALEALEEALK
cd_AtPARP2_633-979     SSNLAPSLIELMKMLFDVETYSAMMEFEINMSEMPLGKLSKHNIQKGFALTEIQRLLT
cd_HsPARP1_662-1007    KSKLPKPVQDLIKMIFDVESMKKAMVEYEIDLQKMPGKLSKRQIQAAYSILSEVQQAVS
cd_GgPARP1_659-1004    KSKLAKPIQDLIKMIFDVESMKKAMVEFEIDLQKMPGKLSKRQIQSAYSILNEVQQAVS
cd_AtPARP1_286-633     QSKLDTRVAKFISLICNVSMMAQHMMIEGYNANKLPLGKISKSTISKGYEVLKRISEVID
cd_HsPARP2_231-577     ESQDLDRVQELIKLICNVQAMEEMMEMKYNTKAPLGLKLTVAQIKAGYQSLKKIEDCIR
                        ..*      .:. : . : : . ** : : : . * . :
cd_AtPARP3_449-801      TTKETGQKAEAMWADFSSRWFLSMHSTRPMR-----LHDVNELADHAASAFETVRDINTA
cd_HsPARP3_182-533     G----PTDGGQSLEELSSHFYTVIPHNFGHSQPP--PINSPELLQAKKDLLVLADIELA
cd_AtPARP2_633-979     ESDPQPTMKESLLVDASNRFFTMPSI-----HPH--IIRDEDDFKSVKMLEALQDIEIA
cd_HsPARP1_662-1007    QGS-----SDSQILDLSNRFYTLIPHDFGMKKPP--LLNNADSVQAKVEMLDNLLDIEVA
cd_GgPARP1_659-1004    DGG-----SESQILDLSNRFYTLIPHDFGMKKPP--LLSNLEYIQAKVQMLDNLLDIEVA
cd_AtPARP1_286-633     RYD-----R-TRLEELSGEFYTVIPHDFGFKKMSQFVIDTPQKCLKQKIEMVEALGEIELA
cd_HsPARP2_231-577     AGQ-----HGRALMEACNEFYTRIPHDFGLRTPP--LIRTOQKELSEKIQLEALGDIEIA
                        : . . : : : : . . . : : : : : : : : : : : : : : : : : : : : : : : : : :
cd_AtPARP3_449-801      SRLIGDMRG----DTLDDPLSDRYKKGCKISVVDKESEDYKMMVKYLETTYEPVKVSDV
cd_HsPARP3_182-533     QALQAVSEQEKTVEEVPHPPLDRDYQLLKCQLQLLDSGAPYKVIQTYLEQTGSNH-----
cd_AtPARP2_633-979     SRIV--GFD----VDSTESLDDKYKKLHCDISPLPHDSEYRLIEKYLNTTHAPT---HT
cd_HsPARP1_662-1007    YSLLRGGSD----DSSKDPIDVNYEKLKTDIKVVDRDSEAEIRKYVKNTHATT---HN
cd_GgPARP1_659-1004    YSLLRGGNE----DGDKPIDINYEKLRTDIKVVDKDSEAEIKQYVKNTHAAT---HN
cd_AtPARP1_286-633     TKLVSVDP-----GLQDDPLYHYQYLNCLGTPVGNDEEFMSVMANYMENTHAKT---HS
cd_HsPARP2_231-577     IKLVKTEL-----QSPEPLDQHYRNLHCALRPLDHESYEFKVISQYLQSTHAPT---HS
                        : . : * . * : : : : : : : : : : : : : : : : : : : : : : :
cd_AtPARP3_449-801      EYGVSVQNVFAVESDAIPS-LDDIK-KLPNKVLLWCGSRSSNLLRHIYKGFPAVCSLVP
cd_HsPARP3_182-533     -RCPTLQHIWKVNQEGEEDRFQAHS-KLGNRKLWHGTNMAVVAAILTSGLRIMPH----
cd_AtPARP2_633-979     EWSLELEEVFALEREGEFDKYAPHREKLGKMLLWHGSRLTNFVGIILNQGLRIAPPEAPA
cd_HsPARP1_662-1007    AYDLEVIDIFKIEREGECQRYKPFK-QLHNRRLWHGSRRTNFAGILSQGLRIAPPEAPV
cd_GgPARP1_659-1004    AYDLKVEIFRIEREGESQRYKPFK-QLHNRQLLWHGSRRTNFAGILSQGLRIAPPEAPV
cd_AtPARP1_286-633     GYTVEIAQLFRASRAVEADRFQQFS-SSKNRMLLWHGSRLTNWAGILSQGLRIAPPEAPV
cd_HsPARP2_231-577     DYTMTLLDLFEVEKDGKEAFRE---DLHNRMLLWHGSRMSNWVILSHGLRIAPPEAPI
                        : . : : . . . * : * * * : : : : : : : : : : : : : : : :
cd_AtPARP3_449-801      PGYMFGRAIVCSDAEAEEARYGF--TAVDRPEGFLVLAVASLGEVEFTFTSPPEDTKTLE
cd_HsPARP3_182-533     SGGRVGKGIYFASENSKSAAGYVIGMKCGAHHVGYMFLGEVALGREHHINTDNP-SLKSP
cd_AtPARP2_633-979     TGYMFGKGIYFADLVSKSAQYCY--TCKKNPVGLMLLSEVALGEIHELTKAKY--MDKPP
cd_HsPARP1_662-1007    TGYMFGKGIYFADMVSKSANYCH--TSQGDPIGLILLGEVALGNMYELKHASH--ISKLP
cd_GgPARP1_659-1004    TGYMFGKGIYFADMVSKSANYCH--TSQADPIGLILLGEVALGNMYELKNASH--ITKLP
cd_AtPARP1_286-633     TGYMFGKGVYFADMFSKSANYCY--ANTGANDGVLLLCEVALGDMNELLYSDY-NADNLP
cd_HsPARP2_231-577     TGYMFGKGIYFADMSSKSANYCF--ASRLKNTGLLLLSEVALGQCNEELLEANP-KAEGLL
                        * . * : : . : : * * * * : : * . * : : *
cd_AtPARP3_449-801      DKKIGVKGLGRKKTETESEHFMR---DDIKVPCGRLPVSEH---KDSPLYNEYAVYDPK
cd_HsPARP3_182-533     PGFDSVIARGHTEPDPTQDTELELDGQQVVPVQGPVPCP--EFSSTSFQSEYLIYQES
cd_AtPARP2_633-979     RGKHSTKGLGKVPQDSEFAKWR---GDVTVPCGKPVSSKV---KASELMYNEYIVYDTA
cd_HsPARP1_662-1007    KGKHSVKGLGKTPDPSANISL---DGVDVPLGTGISSGV---NDTSLLYNEYIVYDIA
cd_GgPARP1_659-1004    KGKHSVKGLGKTAPDPTATTTL---DGVEVPLNGISTGI---NDTCLLYNEYIVYDVA
cd_AtPARP1_286-633     PGKLSKGVGKTA PNPSEAQTLE---DGVVVPLGKPVERSK---SKGMMLYNEYIVYVNE
cd_HsPARP2_231-577     QGKHSTKGLGKMAPSSAHFVTLN---GSTVPLGPASDTGILNPDGYTLNYNEYIVYNPN
                        . . . * : . : * * * . . : . * * : :
cd_AtPARP3_449-801      QTSIRFLVEVKY
cd_HsPARP3_182-533     QCRLRYLLEVHL
cd_AtPARP2_633-979     QVKLQFLLKVRF
cd_HsPARP1_662-1007    QVNLKYLLKLF
cd_GgPARP1_659-1004    QVNLKYLLKLF
cd_AtPARP1_286-633     QIKMRYVIQVKF
cd_HsPARP2_231-577     QVRMRYLLKVQF
                        * : : : : : :
#
# Percent Identity Matrix
#
1: cd_AtPARP3_449-801      100.00  21.96  32.46  28.65  28.65  26.61  26.18
2: cd_HsPARP3_182-533     21.96  100.00  35.03  36.31  36.90  36.01  36.61
3: cd_AtPARP2_633-979     32.46  35.03  100.00  51.18  50.00  42.18  45.10
4: cd_HsPARP1_662-1007    28.65  36.31  51.18  100.00  86.99  46.22  46.49
5: cd_GgPARP1_659-1004    28.65  36.90  50.00  86.99  100.00  45.93  47.37
6: cd_AtPARP1_286-633     26.61  36.01  42.18  46.22  45.93  100.00  50.73
7: cd_HsPARP2_231-577     26.18  36.61  45.10  46.49  47.37  50.73  100.00

```

5.3 Pharmacophore annotations

ID	Annotation	Radius	ID	Annotation	Radius
F1	Don2	1.0	F5	Aro	1.0
F2	Acc2	1.0	F6	PiN	1.4
F3	Don	1.0	F7	PiN	1.4
F4	Acc	1.0	+V1	Excl	1.465

The following 5 loops had to be modeled:

Loop	N-terminal anchor	Loop sequence	C-terminal anchor
1	VLKRI	SEVIDRY	DRTRL
2	GFKKM	SQFVID	TPQKL
3	PSEAQ	TLE	DGVVV
4	TKLLS	VDPGLQ	DDPLY
5	MNELL	YSDYNAD	NLPPG

After the side chains had been built, optimized and fine-tuned, all newly modeled parts were subjected to a combined steepest descent and simulated annealing minimization (i.e. the backbone atoms of aligned residues were kept fixed to avoid potential damage). The resulting half-refined model has been saved and obtained the following quality Z-scores: Then a full unrestrained simulated annealing minimization was run for the entire model. The result has been saved, the corresponding Z-scores are listed below:

combined steepest descent and simulated annealing minimisation			full unrestrained simulated annealing minimisation		
Check type	Quality Z-score	Comment	Check type	Quality Z-score	Comment
Dihedrals	-1.621	Satisfactory	Dihedrals	0.454	Optimal
Packing 1D	-2.007	Poor	Packing 1D	-1.210	Satisfactory
Packing 3D	-3.109	Bad	Packing 3D	-1.686	Satisfactory
Overall	-2.463	Poor	Overall	-1.190	Satisfactory

Since the overall quality Z-score improved to -1.190 during the minimization, this fully refined model has been accepted as the final one for this template and alignment.

5.4.1.4 Report section 6 and 7: Model ranking and hybrid model

Since only a single model has been built, there is obviously no need for a final ranking. The model has an overall quality Z-score of -1.190. Again, with only a single model available, no hybrid model could be built. Instead, the model was simply saved as the final one.

NOTE: A Z-score describes how many standard deviations the model quality is away from the average high-resolution X-ray structure. Negative values indicate that the homology model looks worse than a high-resolution X-ray structure. The overall Z-scores for all models have been calculated as the weighted averages of the individual Z-scores using the formula described in 2.3.3.1

Then a full unrestrained simulated annealing minimization was run for the entire model. The result has been saved, the corresponding Z-scores are listed below:

combined steepest descent and simulated annealing minimisation			full unrestrained simulated annealing minimisation		
Check type	Quality Z-score	Comment	Check type	Quality Z-score	Comment
Dihedrals	-1.751	Satisfactory	Dihedrals	0.581	Optimal
Packing 1D	-1.611	Satisfactory	Packing 1D	-0.720	Good
Packing 3D	-2.972	Poor	Packing 3D	-1.279	Satisfactory
Overall	-2.264	Poor	Overall	-0.791	Good

Since the overall quality Z-score improved to -0.791 during the minimization, this fully refined model has been accepted as the final one for this template and alignment.

5.4.2.4 Report section 6s and 7: Model ranking and hybrid model

Since only a single model has been built, there is obviously no need for a final ranking. The model with an overall quality Z-score of -0.791 has been saved. Again, with only a single model available, no hybrid model could be built. Instead, the model was simply saved as the final one.

NOTE: A Z-score describes how many standard deviations the model quality is away from the average high-resolution X-ray structure. Negative values indicate that the homology model looks worse than a high-resolution X-ray structure. The overall Z-scores for all models have been calculated as the weighted averages of the individual Z-scores using the formula described in 2.3.3.1

5.4.3 YASARA homology modelling report: *At*PARP1 (protein stability)

5.4.3.1 Report section 1: Homology modelling target

```
>NEW_ATPARP1_2PAX
QSKLDTRVAKFISLICNVSMMAQHMMIEIGYNANKLPLGKISKSTISKGYEVLKRISEVID
RYDRTRLEELSGEFYTVIPHDFGFKKMSQFVIDTPQKLKQKIEMVEALGEIELATKLLSV
DPGLQDDPLYHYQQLNCGLTPVGNDSSEEFMSVANMENTHAKTHSGYTVEIAQLFRASR
AVEADRFRQQFSSSKNRMLLWHGSRILTWNAGILSQGLRIAPPEAPVTGYMFGKGVYFADMF
SKSANYCYANTGANDGVLLLCEVALGDMNELLYSDYNADNLPPGKLSTKGVGKTAPNPSE
AQTLEDGVVVPLGKPVERSCSKGMLLYNEYIVYNVEQIKMRYVIQVKFNYKH
```

The target sequence contains 352 residues in 1 molecule.

5.4.3.2 Report section 4: Secondary structure prediction

```
Sequence: QSKLDTRVAKFISLICNVSMMAQHMMIEIGYNANKLPLGKISKSTISKGYEVLKRISEVID
SecStr   : CCCCCHHHHHHHHHHHCHHHHHHHHHHHCCCCCCCCCCCCCHHHHHHHHHHHHHHHHHHHHH
PreHel   : 00123899999999851999999999872000000011112899998999999999975
PreStr   : 000000000000000000000000000000000000000000000000000000000000
PreCoi   : 988761000000001480000000000278888887666761000010000000000014
```

```
Sequence: RYDRTRLEELSGEFYTVIPHDFGFKKMSQFVIDTPQKLKQKIEMVEALGEIELATKLLSV
SecStr   : CCCHHHHHHHHHCCCCCCCCCCCCCCCCCCCCCHHHHHHHHHHHHHHHHHHHHHHHHHHH
PreHel   : 4316579998854333210000000001222100899999999999868999877752
PreStr   : 000000000000001422100000001011000000000000000000000000000000
PreCoi   : 558341000113554246788777766545568900000000000001131000111346
```

```
Sequence: DPGLQDDPLYHYQQLNCGLTPVGNDSSEEFMSVANMENTHAKTHSGYTVEIAQLFRASR
SecStr   : CCCCCCCCCCHHCCCCCEEEEECCCCCHHHHHHHHHHHHHCCCCCCCCCEEEEECEEEEC
PreHel   : 0000000445544321000000000008999999999986531000000121112212102
PreStr   : 00000000000000112367788410000000000000000000012211345643455642
PreCoi   : 999999954444456663322269999100000000013467887787543244322367
```

```
Sequence: AVEADRFRQQFSSSKNRMLLWHGSRILTWNAGILSQGLRIAPPEAPVTGYMFGKGVYFADMF
SecStr   : HHHHHHHHHHCCCCCEEEEECCCCCHHHHHHHHCCCCCCCCCCCCCCCCCEEEEECCCC
PreHel   : 6887678875420000000000000108899875100000000000000000000000002
PreStr   : 0000000000000015899864100000000000000000000001123444567862100
PreCoi   : 31113201245898941001468998910000248998999998776455432136886
```

```
Sequence: SKSANYCYANTGANDGVLLLCEVALGDMNELLYSDYNADNLPPGKLSTKGVGKTAPNPSE
SecStr   : CCCCCCCCCCCCCCEEEEEEEEECCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
PreHel   : 34444332100000000000000110001222110000000000000000000000000000
PreStr   : 0000112210000147999988643112332110001000000001124454111000000
PreCoi   : 444555557899984200000024578544567788788899998875545888899888
```

```
Sequence: AQTLEDGVVVPLGKPVERSCSKGMLLYNEYIVYNVEQIKMRYVIQVKFNYKH
SecStr   : CCCCCCEEECCCCCCCCCCCCCCCCCCCCCEEEEECCCCCEEEEEEEEEEEEEEEEC
PreHel   : 000000000000000000000000000000000000000000000000000000000000
PreStr   : 0343101676300013332111111224589987102457788899887640
PreCoi   : 8556798323689875556788888776410013996542211100112359
```

5.4.3.3 Report section 5: The target sequence profile (excerpt)

A target sequence profile has been created from the following multiple sequence alignment, which is built from related Uniprot sequences. The colour codes are: **negative**, **positive**, **hydrophilic** and **hydrophobic**. The excerpt contains the first 17 of 86 lines.

```
Target : QSKLDTRVAKFISLICNVSMMAQHMMIEIGYNANKLPLGKISKSTISKGYEVLKRRISEVID
PARP2_AR: QSKLDTRVAKFISLICNVSMMAQHMMIEIGYNANKLPLGKISKSTISKGYEVLKRRISEVID
A7Q0E8_V: ETKLEPRIAKFISLICDVSMMKQMMIEIGYNADKLPGLKLSKSTISKGYDVLKRRIADVIS
PARP2_MA: ETKLETRIAQFISLICNISMCKQRMVEIIGNAEKLPGLKLRKATILKGYHVLKRISDVIS
PRP2A_OR: ETKLETRIASFISLICNISMCKQRMVEIIGNSDKLPGLKLSKSTIFKGYDVLKRISNVIS
A2WPQ2_O: ETKLETRIASFISLICNISMCKQRMVEIIGNADKLPGLKLSKSTIFKGYDVLKRISNVIS
PRP2B_OR: ETKLETRTASFISLICDISMMKQMMVEIIGNADKLPGLKLSKSTILKGYDVLKRISNVI.
A9TUE0_P: PSKLNPRLEKEFIELICNVNMMKQMMIEIGYDARKMPLGKLSKSTILKGYEVLKRILAAALD
A9PAR1_P: VTRLDPRIANFISLICDVSMMKQRMELIGNAEKLPGLKLSKSTILKGYDVLRRICENIG
A2WPQ1_O: ETKLETRTASFISLICDISMMKQMMVEIIGNADKLPGLKLSKSTILKGYDVLKRISNVI.
Q4T502_T: ASKLDVKIQSLLLELICDLKAMEECVLEMKFDTRKAPLGLKTPEQIRAGYVALRKIEDCL.
Q24GE4_T: TCKLPKEVISLISLIFDMKMINNQMEIGYDVKKMPLGKLSKENINKAYGMLKQLYEEVE
A5PLJ8_D: PCQLNSKVQSLLLELICDLKAMEECVLEMKFDTKKAPLGLKTAEQIRAGYASLKRIEECL.
A0CA47_P: KSKLHTKIKELVRLIFDMKMINNQMEIGYDAKKMPLGLAASTINKGFDVLKRISEELN
Q566G1_X: QSKLHPLLQSLLOFICDLESMDKAMIEFQIDVKKMPLGKLSKKQIQDALEVLSTLAKRVE
PARP2_HU: ESQDLDRVQELIKLICNVQAMEEMMEMKYNTKKAPLGLKTVAQIKAGYQSLKKIEDCIR
PARP2_MO: ESQDLDRVQELIKLICNVQTEEMMIEMKYDTRAPLGLKTVAQIKAGYQSLKKIEDCIR
```

```
Target : RYDRTRLEELSSEFYTVIPHDFGFKKMSQFVIDTPQKLKQKIEMVEALGETELATKLLSV
PARP2_AR: RYDRTRLEELSSEFYTVIPHDFGFKKMSQFVIDTPQKLKQKIEMVEALGETELATKLLSV
A7Q0E8_V: QSNRKTLEQLSSEFYTVIPHDFGFKKMRDFVIDTPQKLKHKLEMVEALGETEVATKLLKD
PARP2_MA: KADRRHLEQLTGEFYTVIPHDFGFRKMREFIIDTPQKLKAKLEMVEALGETEIAATKLLD
PRP2A_OR: RADRRQLEQLTGEFYTVIPHDFGFRKMREFIIDTPQKLKAKLEMVEALGETEIAATKLLD
A2WPQ2_O: RADRRQLEQLTGEFYTVIPHDFGFRKMREFIIDTPQKLKAKLEMVEALGETEIAATKLLD
PRP2B_OR: ...RTQLEQLTGEFYTVIPHDFGFKMSEFIIIDTPQKLKAKLEMVEALSEIEIAIKLLED
A9TUE0_P: .....SIQELTSEFYTVIPHDFGFKHMQNFIIIDTPQKLKHKLEMVEALGETEVATKLLSN
A9PAR1_P: KSDTEKLEELSSEFYTIIPHDFGFNMREFIIDNHYKLNKCKLEMVEALGETEIAATSLIKD
A2WPQ1_O: ...RTQLEQLTGEFYTVIPHDFGFKMSEFIIIDTPQKLKAKLEMVEALSEIEIAIKLLED
Q4T502_T: .....LLEACNQFYTRIPHDFGLK.....IIQTEQELKDKIALLEALSDIQIIVKMKVKA
Q24GE4_T: .....IEELCNFYYSIIPHDFGFKKMASFILDADKVKKEKLEMTESIQNIQIATKL...
A5PLJ8_D: .....LLDACNQFYTRIPHDFGL.....IIRSEBELKEKITLLETLSDIQIIVKMKVQS
A0CA47_P: .HNTTTLQTLTSEFYTSQIPHDFG.....VINTAQLVKQKLEMLSEIQIQVATKILLEE
Q566G1_X: .....FYTLIPHDFGMKK.....LDNPKIKSKVQMLEDLREIELAYNLIKQ
PARP2_HU: .....ACNEFYTRIPHDFGL.....IRTQKELSEKIQLEALGDIETAIKLVKT
PARP2_MO: AGQHGR..EACNEFYTRIPHDFGL.....IRTEKELSDKVKLEALGDIETAIKLV..
```

```
Target : DPGLQDDPLYHYHQQNLCGLTPVGNDSSEFSMVANYMENTHAKTHSGYTVEIAQLFRASR
PARP2_AR: DPGLQDDPLYHYHQQNLCGLTPVGNDSSEFSMVANYMENTHAKTHSGYTVEIAQLFRASR
A7Q0E8_V: DIGTQEDPLHMHYQRLHCEMIPLVNSEEFSMIAKYMENTHAETHSNYTVDIVQIFRVSR
PARP2_MA: DSSDQDDPLYARYKQLHCDFTPLEADSEYSMIKSYLRNTHGKTHSGYTVDIVQIFKVS
PRP2A_OR: DSTDQDDPLYARYKQLSDFTPLEVGSEYSMIKTYLANTHGKTHSYTVDIVQIFKVS
A2WPQ2_O: DSTDQDDPLYARYKQLSDFTPLEVGSEYSMIKTYLANTHGKTHSYTVDIVQIFKVS
PRP2B_OR: DSSDQDDPLYARYKQFCCDFTPLEVDSEYSMIKTYLANTHGKTHGYTVDIVQIFKVS
A9TUE0_P: DNEDDDPAYTHYKRLNCEMPLDTSDEYALVKQYMEKTHGQTHYGYKLELLNVFKLQR
A9PAR1_P: DIYTQKDPYLYSKYHCLRCELVPLDVVSKEFSMIEKYIRNTGDETH..YRIDIVQIFRASR
A2WPQ1_O: DSSDQDDPLYARYKQFCCDFTPLEVDSEYSM...YLTNTHGKTYTGYTVDIVQIFKVS
Q4T502_T: NEDSDENPLDRQYRALQCRQLDAGCHEYEVIEKYLQSTHAPTHSDYTMVLDIFGVDR
Q24GE4_T: ..GQQINQIQSNYEKLNKCKIEPV..DQVVRKIIDYLNKTHASTHNQYGLTIDEIFEVER
A5PLJ8_D: NVKSDEHPLDRQYHSLNLCQLPLDTSNEYKVIKYLKSTHAPTHDYMMLLDVFAVER
A0CA47_P: QKDDDTNVIDENFKKLGINMQYLDPSKDKVIVKEFVKNTHCPTHKNYDLVDLDFELQK
Q566G1_X: DLEQDVNPLDQHYRQLRTHLELLDTSNDEFARIQQYVKLTHGETHSSYKLEVVSVDVER
PARP2_HU: ELQSPHPLDQHYRNLHLCALRPLDHESYEFKVISQYLQSTHAPTHSDYTMMLLDLFEVEK
PARP2_MO: .....EHPLDQHYRNLHLCALRPLDHESNEFKVISQYLQSTHAPTHKDYTMMLLDVFEVEK
```

Target : AVEADRFQQFSSSKNRMLLWHSRLTNWAGILSQGLRIAPPEAPVTGYMFGKGVYFADMF
 PARP2_AR: AVEADRFQQFSSSKNRMLLWHSRLTNWAGILSQGLRIAPPEAPVTGYMFGKGVYFADMF
 A7Q0E8_V: EGEVERFRKFSSTKNRMLLWHSRLTNWTGILSQGLRIAPPEAPATGYMFGKGVYFADMF
 PARP2_MA: HGETERFQKFASTNRMLLWHSRLSNWAGILSQGLRIAPPEAPVTGYMFGKGVYFADMF
 PRP2A_OR: HGEMERFQKFATAGNRMLLWHSRLTNWAGILSQGLRIAPPEAPVTGYMFGKGVYFADMF
 A2WPQ2_O: HGEMERFQKFATAGNRMLLWHSRLTNWAGILSQGLRIAPPEAPVTGYMFGKGVYFADMF
 PRP2B_OR: LGEMERFQKFASAGNRMLLWHSRLTNWAGILSQGLRIAPPEAPISGFMFGKGVYFADMF
 A9TUE0_P: EGENDRFQNF EKDPNRMLLWHSRLSNWTGILSQGLRIAPPEAPVTGYMFGKGVYFADMF
 A9PAR1_P: EGENERFQKFSQTKNRMLLWHSRLTNWTGILSEGLRIAPPEAP.....GNGLYFGDMF
 A2WPQ1_O: LGEMERFQKFASAGNRMLLWHSRLTNWAGILSQGLRIAPPEAPISGFMFGKGVYFADMF
 Q4T502_T: EGESDSF...SDLPNRTLLWHSRLSNWVGILSQGLRVAPPEAPVTGYMFGKGIYFADMS
 Q24GE4_T: EGENDR.....DIKNKMLLWHSRLTNFVGILSQGLRIAPPEAPVTGYMFGKGVYFADMC
 A5PLJ8_D: EG EKDNFN..SELQNRMLLWHSRLSNWVGILSQGLRVAPAEAPVTGYMFGKGIYFADMS
 A0CA47_P: DQDDNRF.....NRMLLWHSRLTNFVGILSQGLRIAPPEAPVTGYMFGKGVYFADMF
 Q566G1_X: EDERARFEGYT...RQLLWHSRRTNWVGILSQGLRIAPPEAPVTGYMFGKGIYFADMF
 PARP2_HU: DGEKEAFR.....NRMLLWHSRMSNWVGILSHGLRIAPPEAPITGYMFGKGIYFADMS
 PARP2_MO: EGEKEAFR.....NRMLLWHSRLSNWVGILSHGLRVAPPEAPITGYMFGKGIYFADMS

Target : SKSANYCYANTGANDGVLLLLCEVALGDMNELLYSDYNADNLPPGKLS TKGVGKTAPNPSE
 PARP2_AR: SKSANYCYANTGANDGVLLLLCEVALGDMNELLYSDYNADNLPPGKLS TKGVGKTAPNPSE
 A7Q0E8_V: SKSANYSYPCAMTTGVLVLC E VALGDMAELLTANCNADKLP E GKLS TKGIGATAPDPSE
 PARP2_MA: SKSANYCYASEACRSGVLLLLCEVALGDMNELLNADYDANNLPKGKLSKGVGQTAPNMVE
 PRP2A_OR: SKSANYCYASEACRSGVLLLLCEVALGEMNELLNADYDANNLPKGKLS TKGVGQTEPNTAE
 A2WPQ2_O: SKSANYCYASEACRSGVLLLLCEVALGEMNELLNADYDANNLPKGKLS TKGVGQTEPNTAE
 PRP2B_OR: SKSANYCCASEACKSGVMLLCEVALGEMNELLYGDFGADNLPNGKLS TKGVGQTEPNIAE
 A9TUE0_P: SKSANYCCTHANDPIGVLLLSEVALGEMNELLRSDYHANKLPAGKLS TKGVGRTFPDPKE
 A9PAR1_P: SKSAPYCHANWINSDAVLVLC E VALGDM...YGSFN..KLPKGKLSVKVAGTVPDSSQ
 A2WPQ1_O: SKSANYCCASEACKSGVLLLLCEVALGDMNELLYGDFGADNLPNG...GVGQTEPNIAE
 Q4T502_T: SKSANYCFANQSNHVGLLLLLCEVALGDSNELLDADYEANNLPNGKHSTKGLGRTGPDPKN
 Q24GE4_T: SKSANYCFTNKANNTGLMLLCEVALGEMNDKYADYASNLPAGKHSTRGRGKTAPPES
 A5PLJ8_D: SKSANYCFASQKNNQGLLLLSEVALGDSNELLDADYNADQLPSGKHSTKGLGQTAPDPKK
 A0CA47_P: SKSANYCAVTRENNTGLILLCDVALGNTNEKFYSDYANNLPKGHSTWGKGTMPPPAQ
 Q566G1_X: SKSANYCFTSRNQPEGLLLLCEVILGDMHE.....NASPLPPGTHSRKGVGSTQPD PST
 PARP2_HU: SKSANYCFASRLKNTGLLLLSEVALGQCNEELLEANPKAEGLLQKHSTKGLGKMAPSSAH
 PARP2_MO: SKSANYCFASRLKNTGLLLLSEVALGQCNEELLEANPKAQGLLRGKHSTKGMGKMAPSPA

Target : AQTLEDGVVPLGKPVERSCSKGM LLYNEYIVYNVEQIKMRYVIQVKFNYKH
 PARP2_AR: AQTLEDGVVPLGKPVERSCSKGM LLYNEYIVYNVEQIKMRYVIQVKFNYKH
 A7Q0E8_V: AQAFENGIVVPLGKPKLRSDPKGGLLYNEYIVYNVDQIRMRYVVQVTFNFKR
 PARP2_MA: SKVADDGVVPLGEPKQEPSKRGG LLYNEYIVYNVDQIRMRYVLHVNFNFKR
 PRP2A_OR: SKITDDGVVPLGKPKAEP SKRGS LLYNEFIVYNVDQIRMRYVLHVSNFNFKK
 A2WPQ2_O: SKITDDGVVPLGKPKAEP SKRGS LLYNEFIVYNVDQIRM R.....
 PRP2B_OR: SKITDDGMVPLGKPEKVP SRRGSLMYNEYIVYNVDQIRMRYILNVNFNFKR
 A9TUE0_P: YKTLENGVVVVPVQPISSPLSMGCL EYNEYIVYDV SQIRMRYLLQVKFNRYR
 A9PAR1_P: AQTLEDGVLVPLGKPVLPYSQGMWPRNEYIILDVDQIRIRYVVHAKFCYQT
 A2WPQ1_O: SKITDDGMVPLGKPK.....KGS LMYNEYIVYNVDQIRMRYILNVTFNFQR
 Q4T502_T: ALTL..GVTVPMGPGVNTGVGK..LLYNEFVIYNPAQIRMRYLLRIKFNYS
 Q24GE4_T: YVTIYDDVQVPVVGK.....LLYNEFIVYDIRQIK.....
 A5PLJ8_D: SVSL..GVTVPLGPSVKTG.....LLYNEYIVYNPAQIQMKYLLRVQFNFS
 A0CA47_P: NIPF.....PIGKGAPSGVANTS LLYNEFIVYDVAQIRLKYLIKMKWNYK.
 Q566G1_X: YYTSPDGVVYPIGKPK.....LLYNEYIVYDVAQVLQKYLVRVKFLYN.
 PARP2_HU: FVTL..GSTVPLGPASDTG.....YNEYIVYNPNQVRMRYLLKIQFNFLQ
 PARP2_MO: FITL..GSTVPLGPASDTG.....YNEFIVYSFNQVRMRYLLKIQFNFLQ

5.4.3.4 Report section 6: The initial homology models

5.4.3.4.1 Homology Model 1 of 4, based on template 2PAX, alignment variant 1

```

SecStr  : CCCCCHHHHHHHHHHHCHHHHHHHHHHHCCCCCCCCCCCCCHHHHHHHHHHHHHHHHHHHHH
Target  : QSKLDTRVAKFISLICNVSMMAQHMMELIGYNANKLPLGKISKSTISKGYEVLKRISEVID
Match   : :SKL | :I:|I :V: M : M:E: : :K|PLGK|SK I:::Y:|L::|:|:|:
Template: KSKLAKPIQDLIKMIFDVESMKKAMVEFEIDLQKMPLGKLSKRQIQSAYSILNEVQQAVS
SecStr  : CCCCHHHHHHHHHHHHHCHHHHHHHHHHHCCCCCCTTTTCHHHHHHHHHHHHHHHHHHHHH

SecStr  : CCCHHHHHHHHHHHCCCCCCCCCCCCCCCCCCCCCHHHHHHHHHHHHHHHHHHHHHHHHH
Target  : RYDRTRLEELSGEFYTVIPHDFGFKMSQFVIDTPQKLKQKIEMVEALGETELATKLLSV
Match   : ::::| |LS::FYT:IPHDFG:KK : | | : K|M:| L |IE:A :LL
Template: D.SESQILDLSNRFYTLIPHDFGMKKP.....NLEYIQAKVQMLDNLLEIEVAYSLLR.
SecStr  : H.HHHHHHHHHHHHHHHHCCCTTTTCC.....HHHHHHHHHHHHHHHHHHHHHHHH

SecStr  : CCCCCCCCCCHHHCCCCCEEEEECCCCCHHHHHHHHHHHHHCCCCCCCCCEEEEECEEEEECC
Target  : DPGLQDDPLYHYHQQLNCGLTPVGNDSEEFMSVANYMENTHAKTHSGYTVIEAQLFRASR
Match   : DP| :Y|:L: | V :DSEE ::| :Y::NTHA TH::Y ::|:|FR :R
Template: .....KDPIDINYEKLRTDIKVVDDKDEEAKIIKQYVKNTHAATHNAYDLKVVEIFRIER
SecStr  : .....CHHHHHHHHHHCEEEEEEECCCCCHHHHHHHHHHHHHHCCCTTTTCCEEEEEEEEEEEE

SecStr  : HHHHHHHHHHHCCCCCEEEEEEECCCCCHHHHHHHHHHHCCCCCCCCCCCCCCCCCEEEEECCCC
Target  : AVEADRFQQFSSSKNRMLLWHGSRLLTNWAGILSQGLRIAPPEAPVTGYMFGKGVYFADMV
Match   : E::R|: F:: NR:LLWHGSR TN:AGILSQGLRIAPPEAPVTGYMFGKG|YFADM
Template: EGESQRYKPFKQLHNRQLLWHGSRLLTNFAGILSQGLRIAPPEAPVTGYMFGKGIYFADMV
SecStr  : .....CHHHHHHHHHHCEEEEEEECCCCCHHHHHHHHHHHHHHCCCTTTTCCEEEEEEEEEEEE

SecStr  : CCCCCCCCCCCCCCEEEEEEEEECCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
Target  : SKSANYCYANTGANDGVLLLCEVALGDMNELLYSDYNADNLPPGKLSKGVGKTAPNPSE
Match   : SKSANYC|:: : G:|LL EVALG:M EL :LP GK S:KG:GKTAP:P:
Template: SKSANYCHTSQADPIGLILLGVALGNMYELK.....KLPKGKHSVKGLGKTAPDPTA
SecStr  : TTGGGGCCETTEEEEEEEEEEECCCCCEEECC.....CCTTTTCEEEEECEEEEEETT

SecStr  : CCCCCCEEECCCCCCCCCCCCCCCCCCCCCEEEEEEECCCCCEEEEEEEEEEEEEEECC [PsiPred]
Target  : AQTLEDGVVPLGKPVERSCSKGMLLYNEYIVYNVEQIKMRYVIQVKFNYKH
Match   : : TL GV VPLG: |: : : LLYNEYIVY:V Q|:|Y:|::KFNYK
Template: TTTL..GVEVPLGNIGISTGINDTCLLYNEYIVYDVAQVNLKYLLKLFNYKT
SecStr  : CEEE..TEEECCCEEECCCCCCCCCCCCCCCCCEEEECTTTEEEEEEEEEEEEEEECC [YASARA]

```

In the complete template multiple sequence alignment, 330 of 352 target residues (93.8%) are aligned to template residues. Among these aligned residues, the sequence identity is 48.5% and the sequence similarity is 68.2% ('similar' means that the BLOSUM62 score is > 0). The following 5 loops had to be modelled:

Loop	N-terminal anchor	Loop sequence	C-terminal anchor
1	EVIDR	YDRT	RLEEL
2	EAQTL	ED	GVVVP
3	GFKKM	SQFVID	TPQKL
4	TKLLS	VDPGLQD	DPLYY
5	MNELL	YSDYNAD	NLPPG

After the side chains had been built, optimized and fine-tuned, all newly modelled parts were subjected to a combined steepest descent and simulated annealing minimization (i.e. the backbone atoms of aligned residues were kept fixed to avoid potential damage).

The resulting half-refined model has been saved and obtained the following quality Z-scores:

Then a full unrestrained simulated annealing minimization was run for the entire model. The result has been saved, the corresponding Z-scores are listed below:

combined steepest descent and simulated annealing minimisation			full unrestrained simulated annealing minimisation		
Check type	Quality Z-score	Comment	Check type	Quality Z-score	Comment
Dihedrals	0.885	Optimal	Dihedrals	1.314	Optimal
Packing 1D	-1.142	Satisfactory	Packing 1D	-0.533	Good
Packing 3D	-1.003	Satisfactory	Packing 3D	-1.037	Satisfactory
Overall	-0.783	Good	Overall	-0.499	Good

Since the overall quality Z-score improved to -0.499 during the minimization, this fully refined model has been accepted as the final one for this template and alignment.

After the side chains had been built, optimized and fine-tuned, all newly modelled parts were subjected to a combined steepest descent and simulated annealing minimization (i.e. the backbone atoms of aligned residues were kept fixed to avoid potential damage).

The resulting half-refined model has been saved and obtained the following quality Z-scores:

Then a full unrestrained simulated annealing minimization was run for the entire model. The result has been saved, the corresponding Z-scores are listed below:

combined steepest descent and simulated annealing minimisation			full unrestrained simulated annealing minimisation		
Check type	Quality Z-score	Comment	Check type	Quality Z-score	Comment
Dihedrals	0.851	Optimal	Dihedrals	1.034	Optimal
Packing 1D	-1.251	Satisfactory	Packing 1D	-0.691	Good
Packing 3D	-1.034	Satisfactory	Packing 3D	-1.159	Satisfactory
Overall	-0.845	Good	Overall	-0.658	Good

Since the overall quality Z-score improved to -0.658 during the minimization, this fully refined model has been accepted as the final one for this template and alignment.

After the side chains had been built, optimized and fine-tuned, all newly modelled parts were subjected to a combined steepest descent and simulated annealing minimization (i.e. the backbone atoms of aligned residues were kept fixed to avoid potential damage).

The resulting half-refined model has been saved and obtained the following quality Z-scores:

Then a full unrestrained simulated annealing minimization was run for the entire model. The result has been saved, the corresponding Z-scores are listed below:

combined steepest descent and simulated annealing minimisation			full unrestrained simulated annealing minimisation		
Check type	Quality Z-score	Comment	Check type	Quality Z-score	Comment
Dihedrals	0.777	Optimal	Dihedrals	1.210	Optimal
Packing 1D	-1.351	Satisfactory	Packing 1D	-0.818	Good
Packing 3D	-1.055	Satisfactory	Packing 3D	-1.061	Satisfactory
Overall	-0.905	Good	Overall	-0.637	Good

Since the overall quality Z-score improved to -0.637 during the minimization, this fully refined model has been accepted as the final one for this template and alignment.

5.4.3.4.4 Homology Model 4 of 4, based on template 2PAX, alignment variant 4

```

SecStr  : CCCCCHHHHHHHHHHHCHHHHHHHHHHHCCCCCCCCCCCCCHHHHHHHHHHHHHHHHHHHHH
Target  : QSKLDTRVAKFISLICNVSMMAQHMMIEIGYNANKLPLGKISKSTISKGYEVLKRISEVID
Match   : :SKL | :I:|I :V: M : M:E: : :K|PLGK|SK I:::Y:|L::|:|:|:
Template: KSKLAKPIQDLIKMIFDVESMKKAMVEFEIDLQKMPPLGKLSKRQIQSAYSILNEVQQAVS
SecStr  : CCCCCHHHHHHHHHHHCHHHHHHHHHHHCCCCCCTTTTCHHHHHHHHHHHHHHHHHHHHH

SecStr  : CCCHHHHHHHHHHHCCCCCCCCCCCCCCCCCCCCCHHHHHHHHHHHHHHHHHHHHHHHHH
Target  : RYDRTRLEELSGEFTYTVIPHDFGFKMSQFVIDTPQKLKQKIEMVEALGETELATKLLSV
Match   : ::::| |LS::FYT:IPHDFG:KK : | |: K||M:| L |IE:A :LL
Template: D.SESQILDLSNRFYTLIPHDFGMKPP.....NLEYIQAKVQMLDNLDDIEVAYSLLR.
SecStr  : H.HHHHHHHHHHHHHHHHHCCCTTTTCC.....HHHHHHHHHHHHHHHHHHHHHHHH

SecStr  : CCCCCCCCCHHHCCCCCEEEEECCCCCHHHHHHHHHHHHHCCCCCCCCCEEEEECEEEEECC
Target  : DPGLQDDPLYHYHQQLNCGLTPVGNDSSEEFMSVANMENTHAKTHSGYTVIEAQLFRASR
Match   : DP| :Y:|L: | V :DSEE ::| :Y::NTHA TH::Y ::|:|FR :R
Template: .....KDPIDINYEKLRTDIKVVDKDSEEAKI IKQYVKNTHAATHNAYDLKVVEIFRIER
SecStr  : .....CHHHHHHHHHHCEEEEECCCCCHHHHHHHHHHHHHHHHHHHHHCTTTTCEEEEEEEEE

SecStr  : HHHHHHHHHHHCCCCCEEEEECCCCCHHHHHHHHHHHCCCCCCCCCCCCCCCCCEEEEECCCC
Target  : AVEADRFQQFSSSKNRMLLWHGSRLLTNWAGILSQGLRIAPPEAPVTGYMFGKGVYFADM
Match   : E::R|: F:: NR:LLWHGSR TN:AGILSQGLRIAPPEAPVTGYMFGK|YFADM
Template: EGESQRYKPFKQLHNRQLLWHGSRLLTNWAGILSQGLRIAPPEAPVTGYMFGKGIYFADMV
SecStr  : CHHHHHHHHGGGGGCEEEEECCCCCHHHHHHHHHHHCCCCCTTTTGGGGTTTTTCEEEEEECTT

SecStr  : CCCCCCCCCCCCCCEEEEEEEEECCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
Target  : SKSANYCYANTGANDGVLLICEVALGDMNELLYSDYNADNLPPGKLSKGVGKTAPNPSE
Match   : SKSANYC|:: : G:|LL EVALG:M EL :LP GK S:KG:GKTAP:P:
Template: SKSANYCHTSQADPIGLILLGEVALGNMYELK.....KLPKGKHSVKGLGKTAPDPTA
SecStr  : TTGGGGGCCETTEEEEEEEEEEEEECCCCCEEEEC.....CCTTTTCEEEEECEEEEEETT

SecStr  : CCCCCCEEECCCCCCCCCCCCCCCCCCCCCEEEEECCCCCEEEEEEEEEEEEEEECC [PsiPred]
Target  : AQTLEDGVVPLGKPVERSCSKGMLLYNEYIVYNVEQIKMRYVIQVKFNYKH
Match   : : DGV VPLG: |: : : LLYNEYIVY:V Q|:|Y:|::KFNYK
Template: TT.TLDGVEVPLGNISTGINDTCLLYNEYIVYDVAQVNLKYLKLFNYKT
SecStr  : CE.EETTEEECCCCCEEEEECCCCCCCCCCCCCEEEECTTTEEEEEEEEEEEEEEECC [YASARA]

```

In the complete template multiple sequence alignment, 331 of 352 target residues (94.0%) are aligned to template residues. Among these aligned residues, the sequence identity is 48.0% and the sequence similarity is 67.7% ('similar' means that the BLOSUM62 score is > 0).

The following 5 loops had to be modelled:

Loop	N-terminal anchor	Loop sequence	C-terminal anchor
1	EVIDR	YDRT	RLEEL
2	PNPSE	AQT	LEDGV
3	GFKKM	SQFVID	TPQKL
4	TKLLS	VDPGLQD	DPLYY
5	MNELL	YSDYNAD	NLPPG

After the side chains had been built, optimized and fine-tuned, all newly modelled parts were subjected to a combined steepest descent and simulated annealing minimization (i.e. the backbone atoms of aligned residues were kept fixed to avoid potential damage).

The resulting half-refined model has been saved and obtained the following quality Z-scores:

Then a full unrestrained simulated annealing minimization was run for the entire model. The result has been saved, the corresponding Z-scores are listed below:

combined steepest descent and simulated annealing minimisation			full unrestrained simulated annealing minimisation		
Check type	Quality Z-score	Comment	Check type	Quality Z-score	Comment
Dihedrals	0.866	Optimal	Dihedrals	1.193	Optimal
Packing 1D	-1.237	Satisfactory	Packing 1D	-0.690	Good
Packing 3D	-1.005	Satisfactory	Packing 3D	-1.153	Satisfactory
Overall	-0.824	Good	Overall	-0.632	Good

Since the overall quality Z-score improved to -0.632 during the minimization, this fully refined model has been accepted as the final one for this template and alignment.

5.4.3.5 Report sections 7 and 8: Model ranking and the hybrid model

The following table lists the 4 monomeric models, comprising residues 1-352, sorted by their overall quality Z-scores. Finally, YASARA tried to combine the best parts of the 4 models to obtain a hybrid model, hoping to increase the accuracy beyond each of the contributors. The model could not be improved by copying parts from other models; nevertheless it was subjected to a final round of simulated annealing minimization in explicit solvent and obtained the following quality Z-scores:

7 Model ranking:			8 the hybrid model		
Rank	Z-score	Model ID	Check type	Z-score	Comment
1	-0.499	1 (5.4.3.4.1)	Dihedrals	1.210	Good
2	-0.632	4 (5.4.3.4.4)	Packing 1D	-0.306	Good
3	-0.637	3 (5.4.3.4.3)	Packing 3D	-1.042	Good
4	-0.658	2 (5.4.3.4.2)	Overall	-0.428	Good

Since this hybrid model scored better than all previous models, it was saved as the final model.

5.4.4.3 Report section 5: The target sequence profile (excerpt)

Target : QSKLDTRVAKFISLICNVSMMAQHMMIEIGYNANKLPLGKISKSTISKGYEVLKRISEVID
 D7U2A8 : ETKLEPRIAKFISLICDVSMKQMMIEIGYNADKPLGKLSKSTISKGYDVLKRIADVIS
 Q5Z8Q9 : ETKLETRIASFISLICNISMMKQMMVEIGYNSDKLPLGKLSKSTIFKGYDVLKRISNVIS
 A9TUE0 : .SKLNPRLKEFIELICNVNMMKQMMIEIGYDARKMPLGKLSKSTILKGYEVLKRLAAALD
 D8R2J2 : QSKLDSRVAQFVSLICDLKMMRQMMVEIGYDARKMPLGKLSKATILKGYQTLKSIQGVL.
 B9I9Q6 : .TRLDPRIANFISLICDVRMMKQRMMEIGYNAEKLPLGKLSKSTILKGYDVLRRICENIG
 F4P043 : ESKLHPSVKELMELCFNMDMMNLQMMIEIGYDTKKMPLGKLSKANIHKGYEVLKRLSDVIQ
 UPI00018 : ESVLHKSLLQDVMTLIFDITWEESVKEMKFDIKKSPGKLTKKQITAGYEALKAVETCID
 Q4T502 : .SKLDVKIQSLELICDLKAMEECVLEMKFDTRKAPLGKLTPEQIRAGYVALRKIEDCL.
 B5X3M4 : .SKLDVKVQSLELICDIKAMEECVLEMKFDTRKAPLGKLTPEQIRAGYSALKKIEECVK
 C5KI71 : .TKLDEQLYGLIKMICDRQLMVDHMRASGVVNVKMPGKISEDMIKAGYBALQAIIEEEL.
 B3RJY6 : .SQLPSAIDLIKLIQDVQAMKAALIEFEIDLKMPGKLSKQIEDAYQVGLGNLQDL..
 C4Q0U1 : .KLHPALQSLLKFICDVKSMEKTMAEFELDLRKMPLGKLSNQIHEAYDVLNSLSQLI.
 E2RHY6 : ESQDLDRVQELIELICNVQAMEETMVEKMYDTKKAPLGKLTVAQIKAGYQSLKIEDCIR
 UPI00017 : .KLPEPVQRLIRLLFDVSMKVMYEFELDLQKMPGKLSRNQLQQAYTTLNEINSMID
 Q566G1 : QSKLHPLQSLQFICDLESMDAMIEFQIDVKKMPGKLSKQIQDALEVLSTLAKRVE
 E2AAA6 : KSNLAEPIQNLMLRIFDVAEMKVMLEFEIDMDKMPGKLSKQIEKAYAVLTELQEIL.

Target : RYDRTRLEELSSEGFYTVIIPHDFGFKKMSQFVIDTPQKIKQKIEMVVEALGEIEIATKLLSV
 D7U2A8 : QSNRKTLEQLSSEGFYTVIIPHDFGFKKMRDFVIDTPQKIKKHEMVEALGEIEVATKLLKD
 Q5Z8Q9 : RADRRQLEQLTSEGFYTVIIPHDFGFKKREFIIDTPQKIKAKLEMVEALGEIEIATKLLSD
 A9TUE0 :ELTSEFYTVIIPHDFGFKHMNFIIIDTPQKIKKHEMVEALGEIEIATKLLSN
 D8R2J2 :LLDLSSEFYTVIIPHDFGFQNIQQTINTIEKLLKHEMVEALGEIEIAAQVLE.
 B9I9Q6 : KSDTEKLEELSSEFYTVIIPHDFGFNKMREFIIDNHYKLLKHEMVEALGEIEIATSLIKD
 F4P043 : AS.....LSSEFYTVIIPHEFGM.....IQTLMLKDKLSMVEALTDIQIATSI..
 UPI00018 : KVDKQELIEACSQFYTRIPHDFGMK.....LIETTDQLKEKLELLEALNEIQVAISIVED
 Q4T502 :LEACNQFYTRIPHDFGLK.....IIQTEQELKDKIALLEALSIDIQIAVKMVKA
 B5X3M4 : RKGSSR.QEACNQFYTRIPHDFGL.....IIRSEELKEKIALLEALSIDIQIAVKMVQS
 C5KI71 :LLDLSGRFYTVVPHDFGFKKMYFIIIDSEEVKQKMQLEEDLQDM.....
 B3RJY6 : ...RTKIVDATNKFYTVIIPHDFGL.....ILDDPKLIQAKTSMDDLLDIAYVNLIKT
 C4Q0U1 : ..DRTQILSESTRFYTVIIPHDFGFK.....LDNKKIITKKIRMLLEDLLEIELAYKMLQT
 E2RHY6 :ACNEFYTVIIPHDFGL.....IRTEKELSDKVQLEALGDIEIAIKLVKT
 UPI00017 :NKFYSLIIPHDFGI.....IDSKEILNSKLEMIIGSLMEIQIAYSM...
 Q566G1 : K.....FYTVIIPHDFGMK.....LDNPKIKSKVQMLEDLREIELAYNILKQ
 E2AAA6 : ...HTTLIDASNRFYTVIIPHNFGI.....ILESSEIKNKCDMLDALLEMEIAYNLLRD

Target : DPGLQDDPLYHYQQNLNCGLTPVGNDSSEEFMSVANYMENTHAKTHSGYTVIEIAQLFRASR
 D7U2A8 : DIGTQEDPLHMHYQRLHCEMIPLEVNSEEFMSIAKYMENTHAETHSNYTVDIVQIFRVSR
 Q5Z8Q9 : DSTDQDDPLYARYQLSCDFTPELVGSEEFYSMIKTYLANTHGKTHTSYTVDVQIFKVS
 A9TUE0 : DNDEDDPAYTHYKRLNCEMEPLDTSDEYALVKQYMEKTHGQTHYGYKLELLNVFKLQR
 D8R2J2 :DDPAFAHYKRLKCKLEPLDQSGEEFKMIQEYLNKTHGQTHRSYDLILQDVFVQR
 B9I9Q6 : DIYTQEDPLYSKYHCLRCELVPLDVVSKEEFSMIEKYIRNTGDETH..YRIDVVQIFRASR
 F4P043 :PMDVNYRSLMCNLVVDRTSDTFKMCVCDYTKLTHGKTHSSYALEVLDVFDVER
 UPI00018 : ENDEND.PIDLNYKSLNCDLTPDRSDDQFKIVKKYVSNTHGSTHTSYTSLVEDVFTVNR
 Q4T502 : NEDSDENPLDRQYRALQCRLQPLDAGCHEYEVIEKYLQSTHAPTHSDYTMVSLDIFGVDR
 B5X3M4 : SAYGDEHPLDRQYNALQCQLQPLSSCSQEQYQVIERYLQTHAPTHSDFNMTVLDIFSVDR
 C5KI71 :NPVDMQYQRLHCDLEALTPEDEEFKMIEKYMLNTHASTHNDFTAKPSAIFRACK
 B3RJY6 : AKDSGKDPVDTHYESLKTDLDDYGSDEFEMVQKYTKNTHASTHSSYTLVKEVFKVNR
 C4Q0U1 : KGDSKRNPDEHYEQLHTKLEPLDSNCEYKLLIDYVRETHGATHQYTLVNLIFEVHR
 E2RHY6 : ALQSPHPLDQHYRKLHLCALHPLDHESHFKVISQYLQSTHAPTHKDYTMLLDVFVEVEK
 UPI00017 :PLDTHYMKLNCAIDVLHSDMNEFNIQQYIMNTHAETHSSYSLNIKDVFKVVR
 Q566G1 : DLEQDVNPLDQHYRQLRTHLELLDTSNDEFARIQQYVVKLTHGETHSSYKLEVVSVDVER
 E2AAA6 : TTDGKQNPDSHYKQLKTDIEILNKSSEEFKMIKDYVQNTHAATHQYKLEIEEVFVVKR

Target : AVEADRFQQFSSSKNRMLLWHGSRLTNWAGILSQGLRIAPPEAPVTGYMFGKGVYFADMF
D7U2A8 : EGEVERFRKFSSTKNRMLLWHGSRLTNWTGILSQGLRIAPPEAPATGYMFGKGVYFADMF
Q5Z8Q9 : HGEMERFQKFATAGNRMLLWHGSRLTNWAGILSQGLRIAPPEAPVTGYMFGKGVYFADMF
A9TUE0 : EGENDRFQNF EKDPNRMLLWHGSRLSNWTGILSQGLRIAPPEAPVTGYMFGKGVYFADMF
D8R2J2 : DEEDAGFRSFSQTPNRMLLWHGSRLTNWTGILSQGLRIAPPEAPSTGYMFGKGVYFADMF
B9I9Q6 : EGENERFKKFSQTKNRMLLWHGSRLTNWTGILSEGLRIAPPEAP.....GNGLYFGDMF
F4P043 : FGESDRYIESSSHK.RMLLWHGSRLTNFVVGILSQGLRIAPPEAPSTGYMFGKGVYFADMF
UPI00018: EVDSARF.....NRTLWHGSRLTNWCGILKQGLRIAPPEAPVTGYMFGKGVYFADMF
Q4T502 : EGESDSF...SDLPNRTLWHGSRLSNWVGILSQGLRVAPPEAPVTGYMFGKGIYFADMS
B5X3M4 : EGEKNGF...SKLHNRMLLWHGSRLSNWVGILSQGLRVAPPEAPVTGYMFGKGIYFADMS
C5KI71 : ASEED.....KDRMLLWHGSRLTNWCGILSSGLRIAPPEAPVTGYMFGKGLYFADSF
B3RJY6 : HGEEGRYEDYKDFHNRMLLWHGSRVTNFVVGILSQGLRIAPPEAPVTGYMFGKGVYFAGLK
C4Q0U1 : DGEDSRFAKC...NKQLLWHGSRQTNWVGILSQGLRIAPPDAPVTGYMFGKGIYFADIV
E2RHY6 : EGEKEAFR.....NRMLLWHGSRLSNWVGILSHGLRIAPPEAPVTGYMFGKGIYFADMS
UPI00017: SGEEKRFKPFKLLHNRKLLWHGSRLTNFAAAILSQGLRIAPKEAPVTGYMFGKGIYFADMF
Q566G1 : EDERARFEGY...RQLLWHGSRRTNWVGILSQGLRIAPPEAPVTGYMFGKGIYFADMF
E2AAA6 : QGEEQRFKPFKLLPNRKLWHGSRTTNFAGILSQGLRIAPPEAPVTGYMFGKGIYFADMF

Target : SKSANYCYANTGANDGVLLLLCEVALGDMNELLYSDYNADNLPPGKLS TKGVGKTAPNPSE
D7U2A8 : SKSANYSYPCAMTTGVVLVCEVALGDMAELLTANCNADKLPEGKLS TKGIGATAPDPSE
Q5Z8Q9 : SKSANYCYASEACRSGVLLLLCEVALGEMNELLNADYDANNLPKGKLS TKGVGQTEPNTAE
A9TUE0 : SKSANYCCTHANDPIGVLLLLSEVALGGMNELLRSDYHANKLPAGKLS TKGVGRTFPDPKE
D8R2J2 : SKSANYCFTTSQNPRGVLLLLCEVALGQMNELYQADYNANRLPPGKLS TKGLGRSVPNSSQ
B9I9Q6 : SKSAPYCHANWINSDAVLVCEVALGDM...YGSFN..KLPKGKLSVKVAGGTVPDSSQ
F4P043 : SKSANYCFTNSRNTGILLCEVALGKTNLDLVQSDYHAD.....STKGIGRNYPDPKQ
UPI00018: SKSANYCWTNSRQPIGILLCEVALGDCNELTSGDYHADKLPKGKHS TKGLGTEPNPKQ
Q4T502 : SKSANYCFANQSNHVGLLLLLCEVALGDSNELLDADYEANNLPNGKHS TKGLGRTGDPKN
B5X3M4 : SKSANYCFANQRNKTGLLLLLSEVALGDSNELLAADYKAAKLLAGKHS TKGLGQTS PDPKN
C5KI71 : SKSANYCFATQKNNRGLMLLCEVALGRSREYTEAD....LGK GK..TKGVGRSGPDPEE
B3RJY6 : ILSANYCNTNSGSPTGLLLLLCEVALGNMHELKQSKY...LPK DTHSTKGLG GTAPNPSQ
C4Q0U1 : SKSANYCFTTQSQPEGLLLLLCEVILGDMNECLQA....DLPPKYHSRK GIGSVTPDPST
E2RHY6 : SKSANYCFASRVKDIGLLLLSEVALGQCNELEANPEAERLLQKHS TKGLGKTAPSPAS
UPI00017: SKSANYCMASHGNNTGLLLLLCEVALGNMDEYKASEY...KLPPGKHS CMGIGRTKPNPAE
Q566G1 : SKSANYCFTSRNQPEGLLLLLCEVILGDMHE.....NASPLPPGTHSRKGVGSTQ PDPST
E2AAA6 : SKSANYCCTHSQSPTGLLLLLCEVALGNMHERYKADY..EKL PKGKHS TLGRGQTE PDPKD

Target : AQTLEDGVVPLGKPVERSCSKGM LLYNEYIVYNVEQIKMRYVIQVKFNYKH
D7U2A8 : AQAFENGIVVPLGKPKLRSDPKGGLLYNEYIVYNVDQIRMRYVVQVTFNFK.
Q5Z8Q9 : SKITDDGVVPLGKPKAEPKSRGSLLYNEFIVYNVDQIRMRYVLHVSNFK.
A9TUE0 : YKTLENGVVVPGQPISSPLSMGCLEYN EYIVYDVSQIRMRYLLQVKFNYRY
D8R2J2 : FKTLDPGVVPLGKPKVSPNSNTSLEYNE YIVYDTKQIRMRYVLQVDFQYK.
B9I9Q6 : AQVLEDGVLVPLGKPVLPYSQGMWPRNE YIILDVDQIRIRYVHVHAKFCYQ.
F4P043 : YIKLEDGVVVPAGSNLETKGSNGYLQYNE YIVYRVQIRIRYLVKMNFKYK.
UPI00018: AVTLADGTTVPLGKPKTCTTKPGYLWYNE FIVYDVAQIKPRFLVKLKFNYK.
Q4T502 : ALTL...VTVPMGPGVNTGVGK..LLYNE FVIYNPAQIRMRYLLRIKFNY..
B5X3M4 : AVTL..GVTVPMGPGMKTGVGAG..LLYNE FIVYNPAQTHMRYLLRVQFNY..
C5KI71 : MLVMKDGLKVPLGK.....LLYNE YIVYSTCQVRMKYLVEVDFTF..
B3RJY6 : AITLENGTVVPLGKSSKSKVTNSSLLYNE YIVYDVSQIRMRYLVKMFNYKY
C4Q0U1 : FHTNKDGVVYPIGKPIDSNVPNTTLCYNE YIVYNVSQVKQKYLVRVKFHYK.
E2RHY6 : FITL...TVPLGPASDT.....YNE FIVYSPNQVRMRYLLKVRNF..
UPI00017: SLFIEDKIEVPLGKPISSNINDTSLLYNE FIVYDISQVKLRYLVKVDNFNY
Q566G1 : YYTSPDGVVYPIGK.....LLYNE YIVYDVAQVLQKYLVRVKFLY..
E2AAA6 : VHKLDDGVEVPYGMGV.....LLYNE YIVYDVAQVKVRYLIRMNFKYKY

Only 17 of the 211 complete alignments are shown

5.4.4.4 Report section 6: The initial homology models

5.4.4.4.1 Homology Model 1 of 3, alignment variant 1

```

SecStr : CCCCCHHHHHHHHHHHCCHHHHHHHHHHHCCCCCCCCCCCCCHHHHHHHHHHHHHHHHHHH
Target : QSKLDTRVAKFISLICNVSMMAQHMMELIGYNANKLPLGKISKSTISKGYEVLKRISEVID
Match  : :SKL | :I:|I :V: M : M:E: : :K|PLGK|SK I:::Y:|L::|:|:|:
Template: KSKLAKPIQDLIKMIFDVESMKKAMVEFEIDLQKMPGKLSKRQIQSAYSILNEVQQAVS
SecStr : CCCCCHHHHHHHHHHHHCCHHHHHHHHHHHCCCTTTTCTTTTCHHHHHHHHHHHHHHHHHHH

SecStr : HCCHHHHHHHHHHHHHCCCCCCCCCCCCCCCCCHHHHHHHHHHHHHHHHHHHHHHHHHHHCC
Target : RYDRTRLEELSSEFYTVIPHDFGFKKMSQFVIDTPQKLKQKIEMVEALGETELATKLLSV
Match  : ::::| |LS::FYT:IPHDFG:KK : | | : K|M:| L |IE:A :LL
Template: D.SESQILDLSNRFYTLIPHDFGMKPP.....NLEYIQAKVQMLDNLLEEVAYSLLR.
SecStr : H.HHHHHHHHHHHHHHHHCCTTTTCC.....HHHHHHHHHHHHHHHHHHHHHHHHHH

SecStr : CCCCCCCCCCHHHHCCCCCEEEEECCCCCHHHHHHHHHHHCCCCCCCCCEEECCCEEEEC
Target : DPGLQDDPLYHYHQQNLNCGLTPVGNDSSEEFMSVANYMENTHAKTHSGYTVETIAQLFRASR
Match  : DP| :Y|:L: | V :DSEE ::| :Y::NTHA TH::Y ::|:|FR :R
Template: .....KDPIDINYEKLRDIDIKVVDKDSEEAKIIKQYVKNTHAATHNAYDLKVVEIFRIER
SecStr : .....CHHHHHHHHHHCEEEEEEECCCCCHHHHHHHHHHHHCCTTTTCCCEEEEEEEEEEE

SecStr : HHHHHHHHHHCCCCCEEECCCCCCCCCHHHHHHCCCCCCCCCCCCCEEECEEEEECCCC
Target : AVEADRFQQFSSSKNRMLLWHGSRLLTNWAGILSQGLRIAPPEAPVTGYMFGKGVYFADMV
Match  : E::R|: F:: NR:LLWHGSR TN:AGILSQGLRIAPPEAPVTGYMFGKG|YFADM
Template: EGESQRYKPFKQLHNRQLLWHGSRLLTNWAGILSQGLRIAPPEAPVTGYMFGKGIYFADMV
SecStr : CHHHHHHHGGGGGCEEEEEEECCCCCHHHHHHHHHHCCTTTTGGGGTTTTCCEEEEECHH

SecStr : CCCCCCCCCCCCCCEEEEEEEEEEECCCHHHCCCCCCCCCCCCCCCCCCCCCCCCCCCC
Target : SKSANYCYANTGANDGVLLLCHEVALGDMNELLYSDYNADNLPPGKLSKGVGKTAPNPSE
Match  : SKSANYC|:: : G:|LL EVALG:M EL :LP GK S:KG:GKTAP:P:
Template: SKSANYCHTSQADPIGLILLGHEVALGNMYELK.....KLPKGKHSVKGLGKTAPDPTA
SecStr : HHHHHCCCCETTEEEEEEEEEEECCCCCEEEEC.....CCTTTTCEEEEECEEEEGGG

SecStr : CEECCCCCEEECCCCCCCCCCCCCCCCCEEEEECCCCCEEEEEEEEEEEEEEECC [PsiPred]
Target : AQTLEDGVVPLGKPVERSCSKGMLLYNEYIVYNVEQIKMRYVIQVKFNYKH
Match  : : TL GV VPLG: |: : : LLYNEYIVY:V Q|:|Y:|::KFNYK
Template: TTTL..GVEVPLNGISTGINDTCLLYNEYIVYDVAQVNLKYLKLFNYKT
SecStr : GEEE..TEEECCCCCEEECCCCCCCCCCCCCEEEECTTTEEEEEEEEEEEEEEECC [YASARA]

```

In the complete template multiple sequence alignment, 330 of 352 target residues (93.8%) are aligned to template residues. Among these aligned residues, the sequence identity is 48.5% and the sequence similarity is 68.2% ('similar' means that the BLOSUM62 score is > 0).

The following 5 loops had to be modelled:

Loop	N-terminal anchor	Loop sequence	C-terminal anchor
1	EVIDR	YD	RTRLE
2	EAQTL	ED	GVVVP
3	GFKKM	SQFVID	TPQKL
4	TKLLS	VDPGLQ	DDPLY
5	MNELL	YSDYNAD	NLPPG

After the side chains had been built, optimized and fine-tuned, all newly modelled parts were subjected to a combined steepest descent and simulated annealing minimization (i.e. the backbone atoms of aligned residues were kept fixed to avoid potential damage).

The resulting half-refined model has been saved and obtained the following quality Z-scores:

Then a full unrestrained simulated annealing minimization was run for the entire model. The result has been saved, the corresponding Z-scores are listed below:

combined steepest descent and simulated annealing minimisation			full unrestrained simulated annealing minimisation		
Check type	Quality Z-score	Comment	Check type	Quality Z-score	Comment
Dihedrals	0.872	Optimal	Dihedrals	1.161	Optimal
Packing 1D	-1.167	Satisfactory	Packing 1D	-0.614	Good
Packing 3D	-0.854	Good	Packing 3D	-1.027	Satisfactory
Overall	-0.726	Good	Overall	-0.549	Good

Since the overall quality Z-score improved to -0.549 during the minimization, this fully refined model has been accepted as the final one for this template and alignment.

5.4.4.4.2 Homology Model 2 of 3, alignment variant 2

```

SecStr  : CCCCCHHHHHHHHHHHCCCCCCCCCCCCCCCCCHHHHHHHHHHHHHHHHHHHHH
Target  : QSKLDTRVAKFISLICNVSMMAQHMMIEIGYNANKLPLGKISKSTISKGYEVLKRISEVID
Match   : :SKL | :I:|I :V: M : M:E: : :K|PLGK|SK I:::Y:|L::|:|:|:
Template: KSKLAKPIQDLIKMIFDVESMKKAMVEFEIDLQKMPLGKLSKRQIQSAYSILNEVQQAVS
SecStr  : CCCCCHHHHHHHHHHHHHCCCCCCCCCCCCCTTTTCTTTTCHHHHHHHHHHHHHHHHHHH

SecStr  : HCCHHHHHHHHHHHHHCCCCCCCCCCCCCCCCCHHHHHHHHHHHHHHHHHHHHHHCC
Target  : RYDRTRLEELSSEFYTVIPHDFGFKMSQFVIDTPQKLIKQKIEMVEALGEIELATKLLSV
Match   : ::::| |LS::FYT:IPHDFG:KK : | |: K||M:| L |IE:A :LL
Template: D.SESQILDLSNRFYTLIPHDFGMKPK.....NLEYIQAKVQMLDNLDDIEVAYSLLR.
SecStr  : H.HHHHHHHHHHHHHHHHCCCTTTTCC.....HHHHHHHHHHHHHHHHHHHHHHHH

SecStr  : CCCCCCCCCCHHHHCCCCCEEEEECCCCCHHHHHHHHHHHHCCCCCCCCCEEECCCEEEEC
Target  : DPGLQDDPLYHYQQLNCGLTPVGNDSSEEFMSVANYMENTHAKTHSGYTVETIAQLFRASR
Match   : DP| :Y:|L: | V :DSEE ::| :Y::NTHA TH::Y ::|:|FR :R
Template: .....KDPIDINYEKLRTDIKVVDDKDSEEAKI IKQYVKNTHAATHNAYDLKVVVEIFRIER
SecStr  : .....CHHHHHHHHHHCEEEEECCCCCHHHHHHHHHHHHCCCTTTTCCCEEEEEEEEEEE

SecStr  : HHHHHHHHHHCCCCCEEECCCCCCCCCHHHHHHCCCCCCCCCCCCCEEECEEEEECCCC
Target  : AVEADRFQQFSSSKNRMLLWHGSR LTNWAGILSQGLRIAPPEAPVTGYMFGKGVYFADMF
Match   : E::R|: F:: NR:LLWHGSR TN:AGILSQGLRIAPPEAPVTGYMFGKG|YFADM
Template: EGESQRYKPFKQLHNRQLLWHGSR TTNFAGILSQGLRIAPPEAPVTGYMFGKGIYFADMV
SecStr  : CHHHHHHHGGGGGCEEEEEEEEECCCCCHHHHHHHHHHCCCTTTTGGGGTTTTCCEEEECHH

SecStr  : CCCCCCCCCCCCCCEEEEEEEEEEECCCHHHCCCCCCCCCCCCCCCCCCCCCCCCCCCC
Target  : SKSANYCYANTGANDGVLLLLCEVALGDMNELLYSDYNADNLPPGKLS TKVGKTAPNPSE
Match   : SKSANYC|:: : G:|LL EVALG:M EL :LP GK S:KG:GKTAP:P:
Template: SKSANYCHTSQADPIGLILLGEVALGNMYELK.....KLPKGKHSVKGLGKTAPDPTA
SecStr  : HHHHHCCCCETTEEEEEEEEEEEEECCCCCEEEEC.....CCTTTTCEEEEECEEEEGGG

SecStr  : CECCCCCEEECCCCCCCCCCCCCCCCCEEEEECCCCCEEEEEEEEEEEEEEECC [PsiPred]
Target  : AQTLEDGVVVPLGKPVERSCSKGM LLYNEYIVYVVEQIKMRYVIQVKFNKYH
Match   : : T DGV VPLG: |: : : LLYNEYIVY:V Q|:|Y:|:KFNKY
Template: TTT.LDGVVPLGNGISTGINDTCLLYNEYIVYDVAQVNLKYLKLFNKYT
SecStr  : GEE.ETTEEECCCEEECCCCCCCCCCCCCEEEECTTTEEEEEEEEEEEEEEECC [YASARA]

```

In the complete template multiple sequence alignment, 331 of 352 target residues (94.0%) are aligned to template residues. Among these aligned residues, the sequence identity is 48.3% and the sequence similarity is 68.0% ('similar' means that the BLOSUM62 score is > 0).

The following 5 loops had to be modelled:

Loop	N-terminal anchor	Loop sequence	C-terminal anchor
1	EVIDR	YD	RTRLE
2	SEAQT	LE	DGVVV
3	GFKKM	SQFVID	TPQKL
4	TKLLS	VDPGLQ	DDPLY
5	MNELL	YSDYNAD	NLPPG

After the side chains had been built, optimized and fine-tuned, all newly modelled parts were subjected to a combined steepest descent and simulated annealing minimization (i.e. the backbone atoms of aligned residues were kept fixed to avoid potential damage).

The resulting half-refined model has been saved and obtained the following quality Z-scores:

Then a full unrestrained simulated annealing minimization was run for the entire model. The result has been saved, the corresponding Z-scores are listed below:

combined steepest descent and simulated annealing minimisation			full unrestrained simulated annealing minimisation		
Check type	Quality Z-score	Comment	Check type	Quality Z-score	Comment
Dihedrals	0.658	Optimal	Dihedrals	0.996	Optimal
Packing 1D	-1.063	Satisfactory	Packing 1D	-0.528	Good
Packing 3D	-0.910	Good	Packing 3D	-1.040	Satisfactory
Overall	-0.742	Good	Overall	-0.545	Good

Since the overall quality Z-score improved to -0.545 during the minimization, this fully refined model has been accepted as the final one for this template and alignment.

5.4.4.4.3 Homology Model 3 of 3, alignment variant 3

```

SecStr  : CCCCCHHHHHHHHHHCCHHHHHHHHHHHCCCCCCCCCCCCCHHHHHHHHHHHHHHHHHHHHH
Target  : QSKLDTRVAKFISLICNVSMMAQHMMIEIGYNANKLPLGKISKSTISKGYEVLKRISEVID
Match   : :SKL | :I:|I :V: M : M:E: : :K|PLGK|SK I:::Y:|L::|:|:|:
Template: KSKLAKPIQDLIKMIFDVESMKKAMVEFEIDLQKMPPLGKLSKRQIQSAYSILNEVQQAVS
SecStr  : CCCCCHHHHHHHHHHHHCCHHHHHHHHHHHCCCTTTTCTTTTCHHHHHHHHHHHHHHHHHHH

SecStr  : HCCHHHHHHHHHHHHHCCCCCCCCCCCCCCCCCHHHHHHHHHHHHHHHHHHHHHHHHHCC
Target  : RYDRTRLEELSGEFTYTVIPHDFGFKMSQFVIDTPQKLKQKIEMVEALGETELATKLLSV
Match   : ::::| |LS::FYT:IPHDFG:KK : | | : K||M:| L |IE:A :LL
Template: D.SESQILDLSNRFYTLIPHDFGMKKP.....NLEYIQAKVQMLDNLDDIEVAYSLLR.
SecStr  : H.HHHHHHHHHHHHHHHHCCTTTTCC.....HHHHHHHHHHHHHHHHHHHHHHHHHH

SecStr  : CCCCCCCCCHHHHCCCCCEEEEECCCCCHHHHHHHHHHHCCCCCCCCCCCCCEEECCCEEEEC
Target  : DPGLQDDPLYHYQQLNCGLTPVGNDSEEFMSVANMENTHAKTHSGYTVIEIAQLFRASR
Match   : DP| :Y|:L: | V :DSEE ::| :Y::NTHA TH::Y ::|:|FR :R
Template: .....KDPIDINYEKLRTDIKVVDKDSEEAKI IKQYVKNTHAATHNAYDLKVVEIFRIER
SecStr  : .....CHHHHHHHHHHCEEEEECCCCCHHHHHHHHHHHHCCTTTTCCCEEEEEEEEEEE

SecStr  : HHHHHHHHHHCCCCCEEECCCCCCCCCHHHHHHCCCCCCCCCCCCCCCCCEEECEEEEEECCC
Target  : AVEADRFQQFSSSKNRMLLWHGSR LTNWAGILSQGLRIAPPEAPVTGYMFGKGVYFADM
Match   : E::R|: F:: NR:LLWHGSR TN:AGILSQGLRIAPPEAPVTGYMFGK|YFADM
Template: EGESQRYKPFKQLHNRQLLWHGSR TTNFAGILSQGLRIAPPEAPVTGYMFGKGIYFADMV
SecStr  : CHHHHHHHGGGGGCEEEEECCCCCHHHHHHHHHHCCTTTTGGGGTTTTTCCEEEEECHH

SecStr  : CCCCCCCCCCCCCCEEEEEEEEEEECCCHHHHCCCCCCCCCCCCCCCCCCCCCCCCCCCC
Target  : SKSANYCYANTGANDGVLLICEVALGDMNELLYSDYNADNLPPGKLS TKGVGKTAPNPSE
Match   : SKSANYC|:: : G:|LL EVALG:M EL :LP GK S:KG:GKTAP:P:
Template: SKSANYCHTSQADPIGLILLGEVALGNMYELK.....KLPKGKHSVKGLGKTAPDPTA
SecStr  : HHHHHCCCCETTEEEEEEEEEEEEECCCCCEEEEC.....CCTTTTCEEEEECEEEEGGG

SecStr  : CECCCCCEEECCCCCCCCCCCCCCCCCEEEEECCCCCEEEEEEEEEEEEEEECC [PsiPred]
Target  : AQTLEDGVVPLGKPVERSCSKGM LLYNEYIVYVQIKMRYVIQVKFNYKH
Match   : : DGV VPLG: |: : : LLYNEYIVY:V Q|:|Y:|::KFNYK
Template: TT.TLDGVEVPLGNIGISTGINDTCLLYNEYIVYDVAQVNLKYLKLFNYKT
SecStr  : GE.EETTEEECCCCCEEECCCCCCCCCCCCCEEEECTTTEEEEEEEEEEEEEEECC [YASARA]

```

In the complete template multiple sequence alignment, 331 of 352 target residues (94.0%) are aligned to template residues. Among these aligned residues, the sequence identity is 48.0% and the sequence similarity is 67.7% ('similar' means that the BLOSUM62 score is > 0).

The following 5 loops had to be modelled:

Loop	N-terminal anchor	Loop sequence	C-terminal anchor
1	EVIDR	YD	RTRLE
2	PSEAQ	TL	EDGVV
3	GFKKM	SQFVID	TPQKL
4	TKLLS	VDPGLQ	DDPLY
5	MNELL	YSDYNAD	NLPPG

After the side chains had been built, optimized and fine-tuned, all newly modelled parts were subjected to a combined steepest descent and simulated annealing minimization (i.e. the backbone atoms of aligned residues were kept fixed to avoid potential damage).

The resulting half-refined model has been saved and obtained the following quality Z-scores:

Then a full unrestrained simulated annealing minimization was run for the entire model. The result has been saved, the corresponding Z-scores are listed below:

combined steepest descent and simulated annealing minimisation			full unrestrained simulated annealing minimisation		
Check type	Quality Z-score	Comment	Check type	Quality Z-score	Comment
Dihedrals	0.849	Optimal	Dihedrals	1.015	Optimal
Packing 1D	-1.046	Satisfactory	Packing 1D	-0.508	Good
Packing 3D	-0.847	Good	Packing 3D	-1.044	Satisfactory
Overall	-0.679	Good	Overall	-0.537	Good

Since the overall quality Z-score improved to -0.537 during the minimization, this fully refined model has been accepted as the final one for this template and alignment.

5.4.4.5 Report sections 7 and 8: Model ranking and the hybrid model

The following table lists the 3 monomeric models, comprising residues 1-352, sorted by their overall quality Z-scores. Finally, YASARA tried to combine the best parts of the 3 models to obtain a hybrid model, hoping to increase the accuracy beyond each of the contributors.

7 Model ranking:		
Rank	Z-score	Model ID
1	-0.537	3 (5.4.4.4.3)
2	-0.545	2 (5.4.4.4.2)
3	-0.549	1 (5.4.4.4.1)

The model could not be improved (with a score of -1.260) by copying parts from other models. The hybrid model was discarded, and the best-ranked model (5.4.4.4.3) was saved as the final model.

5.4.5 Homology model structure evaluation

Protein structure evaluation with RAMPAGE

PDB entry 1UK1

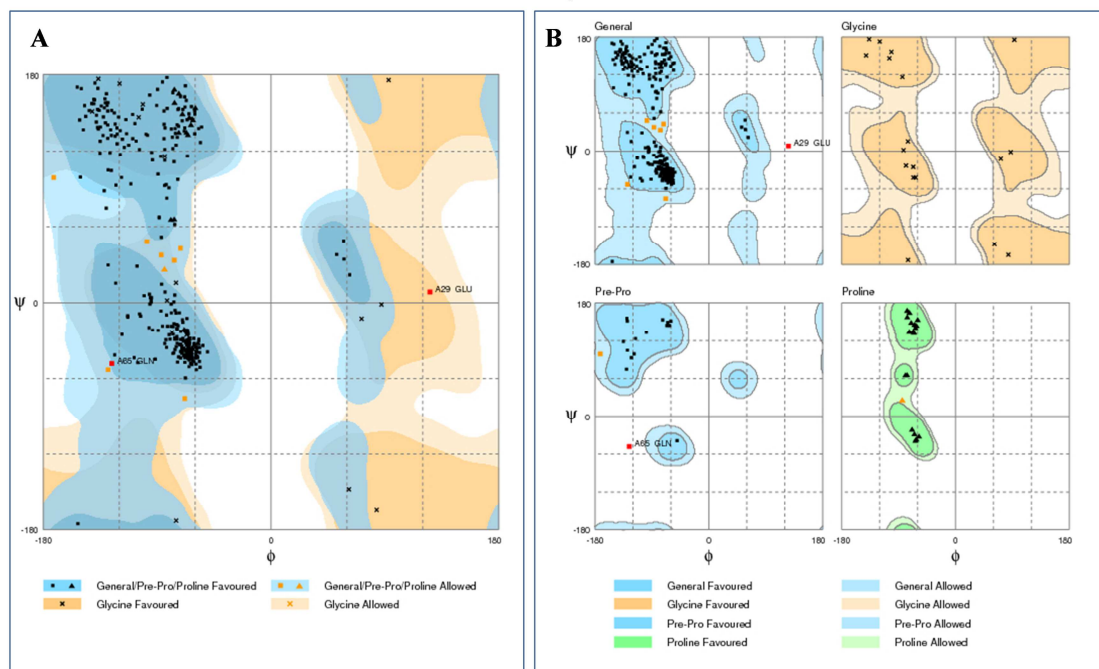


Figure 5.1: RAMPAGE results of PDB entry 1UK1

A: RAMPAGE Ramachandran plot B: RAMPAGE glycine and proline Ramachandran plot

Protein structure evaluation with RAMPAGE

YASARA model of 1UK1

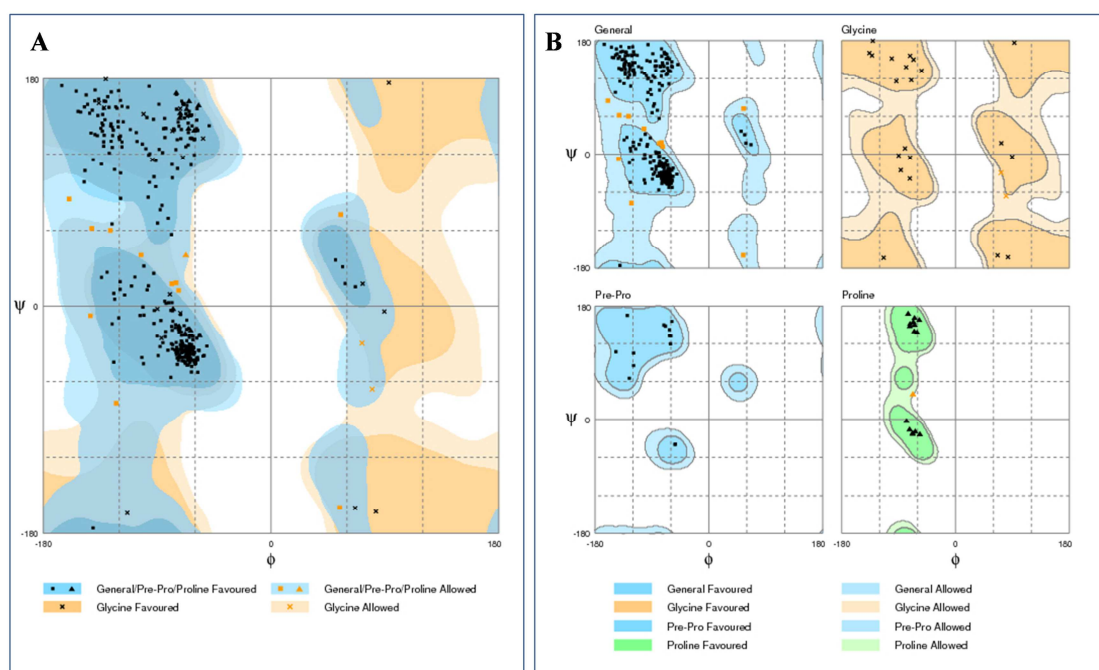


Figure 5.2: RAMPAGE results of YASARA model of 1UK1

A: RAMPAGE Ramachandran plot B: RAMPAGE glycine and proline Ramachandran plot

Protein structure evaluation with RAMPAGE

AtPARP1 homology model

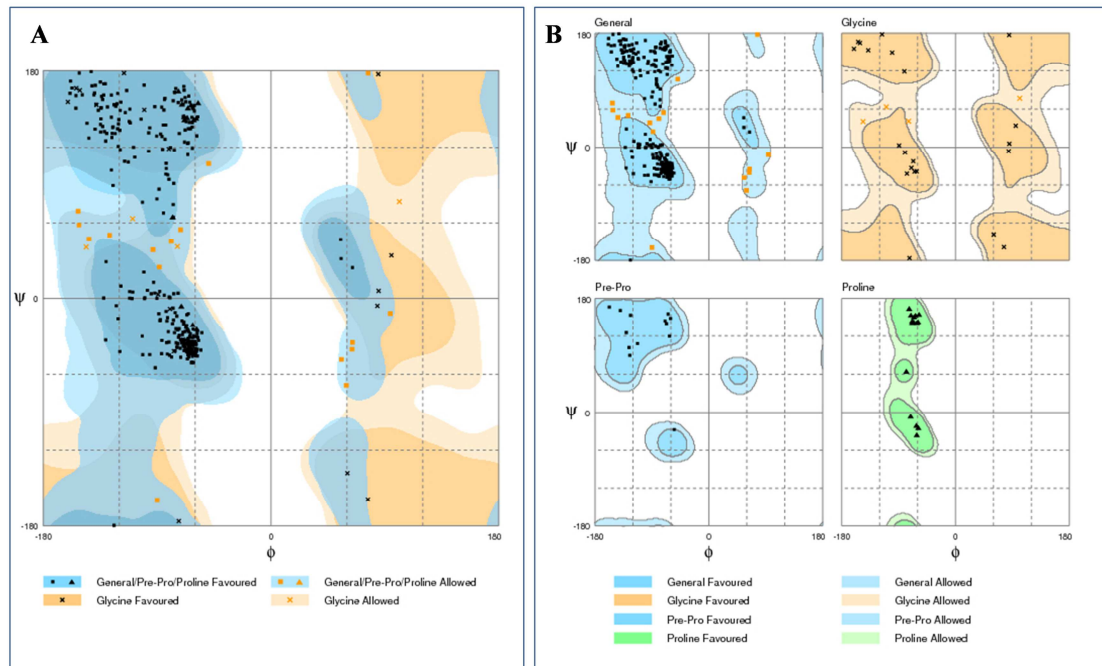


Figure 5.3: RAMPAGE results of homology model of *AtPARP1*

A: RAMPAGE Ramachandran plot B: RAMPAGE glycine and proline Ramachandran plot

Protein structure evaluation with RAMPAGE

AtPARP2 homology model

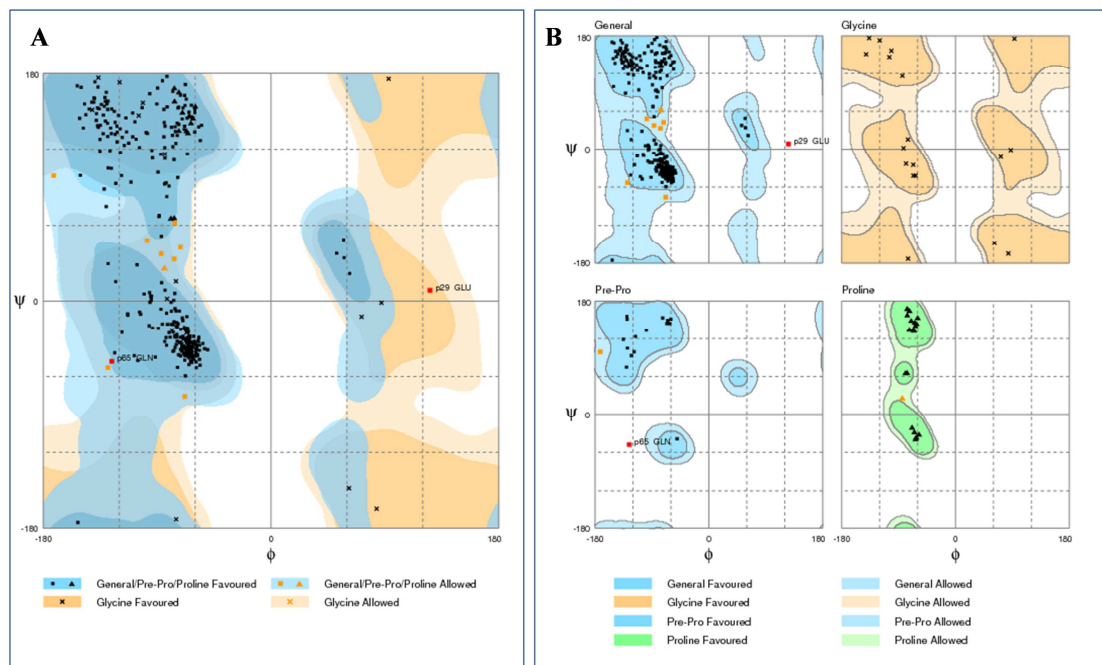


Figure 5.4: RAMPAGE results of homology model of *AtPARP2*

A: RAMPAGE Ramachandran plot B: RAMPAGE glycine and proline Ramachandran plot

5.5 Model refinement

5.5.1 ProSA-web and Errat

	<i>At</i> PARP2			<i>At</i> PARP1		
	Init. model	MD refinement		Init. model	MD refinement	
		Sample mean	Confidence interval *		Sample mean	Confidence interval *
ProSA-web ^c	-8.94	-8.672	-8.764 - -8.579	-8.75	-8.967	-9.023 – 8.911
Errat ^d	94.960	95.653	94.984 - 96.322	98.251	98.173	97.761 – 98.586

* 95% confidence interval; ^c Z-score; ^d overall quality

5.5.2 RAMPAGE

Protein structure evaluation of *At*PARP1 with RAMPAGE

amino acid number	YASARA homology model	YASARA MD refinement																			
		snapshot 0	snapshot 1	snapshot 2	snapshot 3	snapshot 4	snapshot 5	snapshot 6	snapshot 7	snapshot 8	snapshot 9	snapshot 10	snapshot 11	snapshot 12	snapshot 13	snapshot 14	snapshot 15	snapshot 16	snapshot 17	snapshot 18	snapshot 19
Ser 287																					
Lys 288																					
Arg 346																					
Asp 348																					
Gly 368																					
Met 372																					
Asp 378																					
Ser 404																					
Asp 411																					
Gly 424																					
Leu 425																					
Gly 429																					
Asn 430																					
Lys 448																					
Thr 449																					
Gly 452																					
Ala 463																					
Ser 464																					
Val 467																					
Thr 511																					
Lys 517																					
Ala 538																					
Asn 539																					
Leu 557																					
Ser 559																					
Asn 562																					
Asp 564																					
Gln 587																					
Asp 591																					
Leu 597																					
Cys 605																					
Gly 608																					
Met 609																					
Arg 626																					
Asn 634																					
total number of aa in allowed region	20	19	20	14	16	13	10	13	15	11	17	16	13	14	14	16	13	13	12	15	16
total number of outlier	0	1	0	1	0	1	3	1	0	2	1	1	1	2	1	2	1	1	1	0	0

amino acid in allowed region amino acid classified as outlier

Figure 5.5: RAMPAGE - evaluation of *At*PARP1 initial model with MD-refinement

Protein structure evaluation of *At*PARP2 with RAMPAGE

amino acid number	YASARA homology model	YASARA MD refinement																			
		snapshot 0	snapshot 1	snapshot 2	snapshot 3	snapshot 4	snapshot 5	snapshot 6	snapshot 7	snapshot 8	snapshot 9	snapshot 10	snapshot 11	snapshot 12	snapshot 13	snapshot 14	snapshot 15	snapshot 16	snapshot 17	snapshot 18	snapshot 19
Phe 648																					
Phe 660																					
Glu 661																					
Gln 697																					
Pro 698																					
His 720																					
Arg 725																					
Gly 751																					
Phe 752																					
Asp 753																					
Asp 755																					
Glu 758																					
Lys 765																					
Thr 791																					
His 796																					
Trp 799																					
Glu 810																					
Glu 823																					
Lys 824																					
Lys 864																					
Pro 886																					
Tyr 908																					
Pro 942																					
Cys 943																					
Lys 952																					
Ala 953																					
Gln 968																					
Gln 972																					
Phe 979																					
total number of aa in allowed region	9	9	8	6	8	7	10	8	5	8	5	7	4	6	8	7	6	7	6	7	8
total number of outlier	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

amino acid in allowed region amino acid classified as outlier

Figure 5.6: RAMPAGE - evaluation of *At*PARP2 initial model with MD-refinement

5.5.3 Verify 3D

Protein structure evaluation with Verify3D

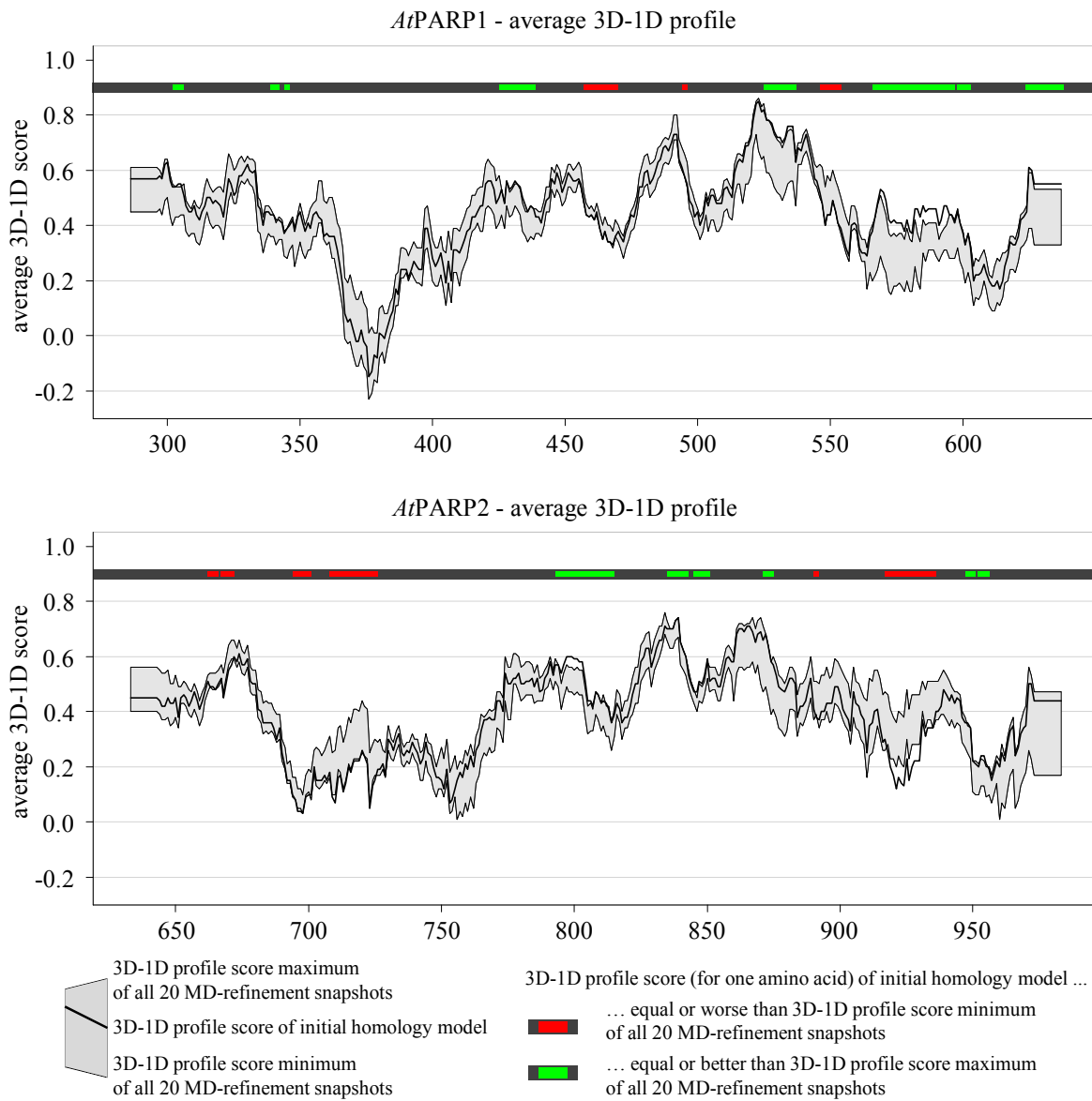


Figure 5.7: Verify 3D - *AtPARP1* and *AtPARP2* initial models and MD-refinement

5.6 Docking

5.6.1 Comparison of docking protocols

Docking program			Novikov loigands docked correctly		NAD ⁺ docked correctly
Refinement	Placement				
MOE	No	Alpha PMI	3	2.1%	No
		Alpha Triangle	49	34.5%	No
		Pharmacophore	125	88.0%	Yes
		Proxy Triangle	66	46.5%	No
		TriangleMatcher	65	45.8 %	No
	Tethered	Alpha PMI	16	11.3%	No
		Alpha Triangle	87	61.3%	No
		Pharmacophore	125	88.0%	Yes
		Proxy Triangle	96	67.6%	No
		TriangleMatcher	94	66.2%	No
	Free	Alpha PMI	15	10.6%	No
		Alpha Triangle	81	57.0%	No
		Pharmacophore	125	88.0%	Yes
		Proxy Triangle	99	69.8%	No
		TriangleMatcher	99	69.8%	No
Scoring function					
GOLD	ASP		72	50.7%	No
	ChemPLP		88	62.0%	No
	ChemScore		89	62.7%	Yes
	GoldScore		87	61.3%	No
Glide	Glide XP		88	62.0%	No
PLANTS (protocol I)	PLP		77	54.2%	No
	PLP95		110	77.5%	Yes
	ChemPLP		139	97.9%	Yes
PLANTS (protocol II)	PLP		88	62.0%	No
	PLP95		112	78.9%	No
	ChemPLP		128	90.1%	Yes

5.6.2 Docking score distributions

Errors of normal approximation for 10 docking runs

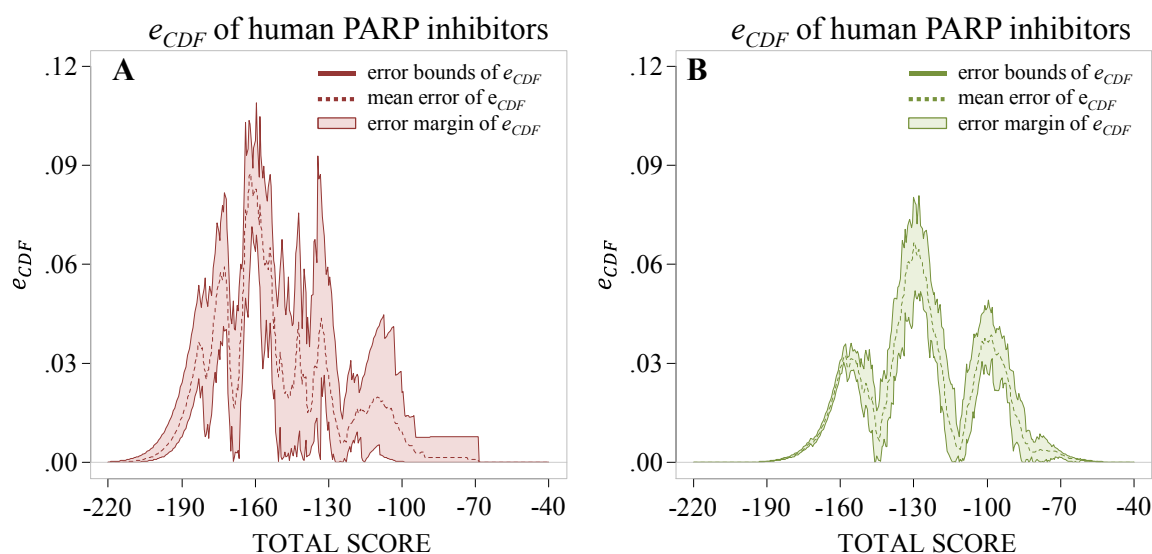


Figure 5.8: Docking score distributions – normal approximation III

Distributions of differences between calculated CDF and estimated CDF (expressed as e_{CDF}) for all 10 independent docking runs with HsPARP1 inhibitors (A) and HsPARP1 decoys (B).

5.6.3 Inference of docking score distributions

docking run	Novikov ligands's TOTAL SCORE		DUD decoys's TOTAL SCORE	
	mean	sd	mean	sd
1	-152.1462	19.1766	-123.5645	19.39006
2	-152.4886	17.3654	-123.3217	19.31003
3	-154.1803	15.4512	-123.2675	19.23195
4	-153.4050	15.7464	-123.4575	19.68753
5	-152.8510	16.3484	-123.1483	19.65138
6	-153.0055	17.1482	-123.3608	19.58277
7	-152.3190	17.7229	-123.5877	19.78588
8	-152.8011	17.1827	-123.0829	19.98123
9	-152.8795	15.4071	-123.5942	19.35270
10	-153.9621	15.6173	-122.8300	19.46453
Mean	-153.0038	16.7166	-123.3215	19.54381
± sd	± 0.6677	± 1.2242	± 0.24849	± 0.2361

5.6.4 Observed and approximated docking scores, differences

docking run	Novikov ligands				DUD decoys			
	docked correctly	threshold missed	type I error obs.	$\Delta(\text{obs, pred})$	docked correctly	threshold passed	type II error obs.	$\Delta(\text{obs, pred})$
1	128	65	50.78%	5.06%	771	13	1.69 %	3.31 %
2	131	67	51.15%	4.70%	755	15	1.99 %	3.01 %
3	128	62	48.44%	7.40%	770	14	1.82 %	3.18 %
4	125	59	47.20%	8.64%	764	12	1.57 %	3.42 %
5	133	66	49.62%	6.22%	745	15	2.01 %	2.99 %
6	128	63	49.22%	6.62%	762	14	1.84 %	3.16 %
7	131	65	49.62%	6.22%	745	16	2.15 %	3.85 %
8	127	62	48.82%	7.02%	765	16	2.09 %	2.91 %
9	130	67	51.54%	4.30%	759	14	1.84 %	3.16 %
10	127	63	49.61%	6.23%	767	14	1.83 %	3.17 %

5.6.5 Number of structures found in pose 1

docking run	<i>Hs</i> PARP1			<i>At</i> PARP1			<i>At</i> PARP2		
	docked correctly	found in pose 1 (%)		docked correctly	found in pose 1 (%)		docked correctly	found in pose 1 (%)	
1	128	100	78.1%	126	76	60.3%	109	90	82.6%
2	131	102	77.9%	117	82	70.1%	108	94	87.0%
3	128	103	80.5%	128	83	64.8%	113	90	79.6%
4	125	96	76.8%	116	83	71.6%	110	94	85.5%
5	133	104	78.2%	122	79	64.8%	113	95	84.1%
6	128	104	81.3%	127	81	63.8%	109	90	82.6%
7	131	103	78.6%	128	72	56.3%	110	91	82.7%
8	127	102	80.3%	121	85	70.2%	111	91	82.0%
9	130	106	81.5%	122	83	68.0%	112	94	83.9%
10	127	104	81.9%	121	79	65.3%	112	93	83.0%
mean	129	102	79.5%	123	80	65.5%	111	92	83.3%

5.6.6 Number of structures docked correctly in docking runs

structures docked correctly in x docking runs	<i>Hs</i> PARP1		<i>At</i> PARP1		<i>At</i> PARP2	
	docked correctly	%	docked correctly	%	docked correctly	%
10	100	70.4%	85	59.9%	98	69.0%
9	16	81.7%	14	69.8%	5	72.5%
8	8	87.3%	10	76.8%	1	73.2%
7	5	90.8%	11	84.5%	1	74.0%
6	2	92.3%	9	90.8%	3	76.1%
5	4	95.1%	3	93.0%	3	78.2%
4	2	96.5%	4	95.8%	3	80.3%
3	0	96.5%	3	97.9%	2	81.7%
2	2	97.9%	0	97.9%	3	83.8%
1	1	98.6%	1	98.6%	10	90.8%
any	2	1.4%	2	1.4%	13	9.2%

5.6.7 Pearson's Chi-squared test results

	<i>At</i> PARP2	<i>At</i> PARP1
Chi-Square Statistic	0.9372	0
Degrees of freedom	1	1
<i>P</i> -value	0.3333	1
Reject H_0	no	no

The Chi-Squared Test with Yates continuity correction for small numbers was used; Pearson's Chi-Squared Test with Yates continuity correction for small numbers²⁷⁰ were used for both *At*PARP1 and *At*PARP2. The null hypothesis H_0 that there is no association between the *P*-values and the mean difference, Δ , of docking scores in *At*PARP2 and *Hs*PARP1 or *At*PARP1 and *Hs*PARP1 was tested. The significance level α of the test was 0.05

5.6.8 PLANTS script for docking

```
# search algorithm settings
aco_ants 20
aco_evap 0.15
aco_sigma 1.0
#
flip_amide_bonds 0
flip_planar_n 1
#
# cluster algorithm settings
cluster_structures 10
cluster_rmsd 2.0
#
#
#protein file
# atPARP2 used, since worked with older nomenclature
protein_file 1UK1_atPARP2_yasara_superpose_receptor_2_molcharge.mol2
#
# mol2 ligand files or multi-mol2 database
ligand_file screen_all_names2_molcharge.mol2
#
#
# binding site definition
bindingsite_center 7.62833 -0.962431 31.4127
bindingsite_radius 12
#
# scoring parameters
scoring_function chemplp
ligand_intra_score lj
#
# adjusting hydrogen bond weights for specific atoms
chemplp_protein_hb_constraint 3235 10 # hb1, Gly863 in HsPARP1
chemplp_protein_hb_constraint 3865 10 # hb3, Ser904 in HsPARP1
chemplp_protein_hb_constraint 3866 10 # hb3, Ser904 in HsPARP1

# set flexible side chains
flexible_protein_side_chain_string SER243# Ser904 in HsPARP1
flexible_protein_side_chain_string GLU103# Glu763 in HsPARP1
flexible_protein_side_chain_string TYR246# Tyr907 in HsPARP1
#
# output directory
output_dir at2_Bionet_selected_molcharge_score2hb_01
```

5.6.9 R script for ROC curves

```

test_pos<-data.frame(c(5,7,8,10,12),rep(1,5))
test_neg<-data.frame(c(3,4,6,7.5,9),rep(0,5))
colnames(test_pos)<-c("data","active")
colnames(test_neg)<-c("data","active")
test_all<-rbind(test_pos,test_neg)

positives<- dim(test_pos)[1]; positives;
negatives<- dim(test_neg)[1]; negatives;

test_all_sorted<-test_all[order(test_all$data,decreasing=T),]

pos_sum<-c(0); neg_sum<-c(0); pos_scaled<-c(0); neg_scaled<-c(0);
for(i in 1:(positives+negatives)) {
  pos_sum<-c(pos_sum,length(which(test_all_sorted[1:i,2]==1)))/positives)
  neg_sum<-c(neg_sum,length(which(test_all_sorted[1:i,2]==0)))/negatives)
  pos_scaled<-c(pos_scaled,length(which(test_all_sorted[1:i,2]==1)))
  neg_scaled<-c(neg_scaled,length(which(test_all_sorted[1:i,2]==0)))
}
length(pos_sum)-1;
length(neg_sum)-1;
pos_sum;
neg_sum;
pos_scaled;
neg_scaled;

plot(c(0,1), type="n",xlim=c(0,1),ylim=c(0,1),
      xlab="inactive fraction",ylab="active fraction",
      main="",#ROC curve",axes=F)

axis(1, at=seq(0,1,by=0.2),labels=rep("",length(seq(0,1,by=0.2))),las=0)
axis(2, at=seq(0,1,by=0.2),labels=rep("",length(seq(0,1,by=0.2))),las=2)
box("plot",col="grey")

# draw lines
abline(h=seq(0,1,.20),col="lightgrey",lty=3)
abline(v=seq(0,1,.20),col="lightgrey",lty=3)
lines(c(0,1),c(0,1),type="l",lty=2)

for(i in 1:length(pos_sum)) {
  lines(c(neg_sum[i], neg_sum[i+1]),c(pos_sum[i],pos_sum[i+1])) }

single_area<-(1/(negatives*positives))
partial_areas<-c()
for(i in 1:negatives) {
  partial_areas<-c(partial_areas,pos_scaled[max(which(neg_scaled<i))])
}
sum(partial_areas);
sum(partial_areas)*single_area;

```

5.6.10 Docking results – confidently docked inhibitors

Analysis of docking results

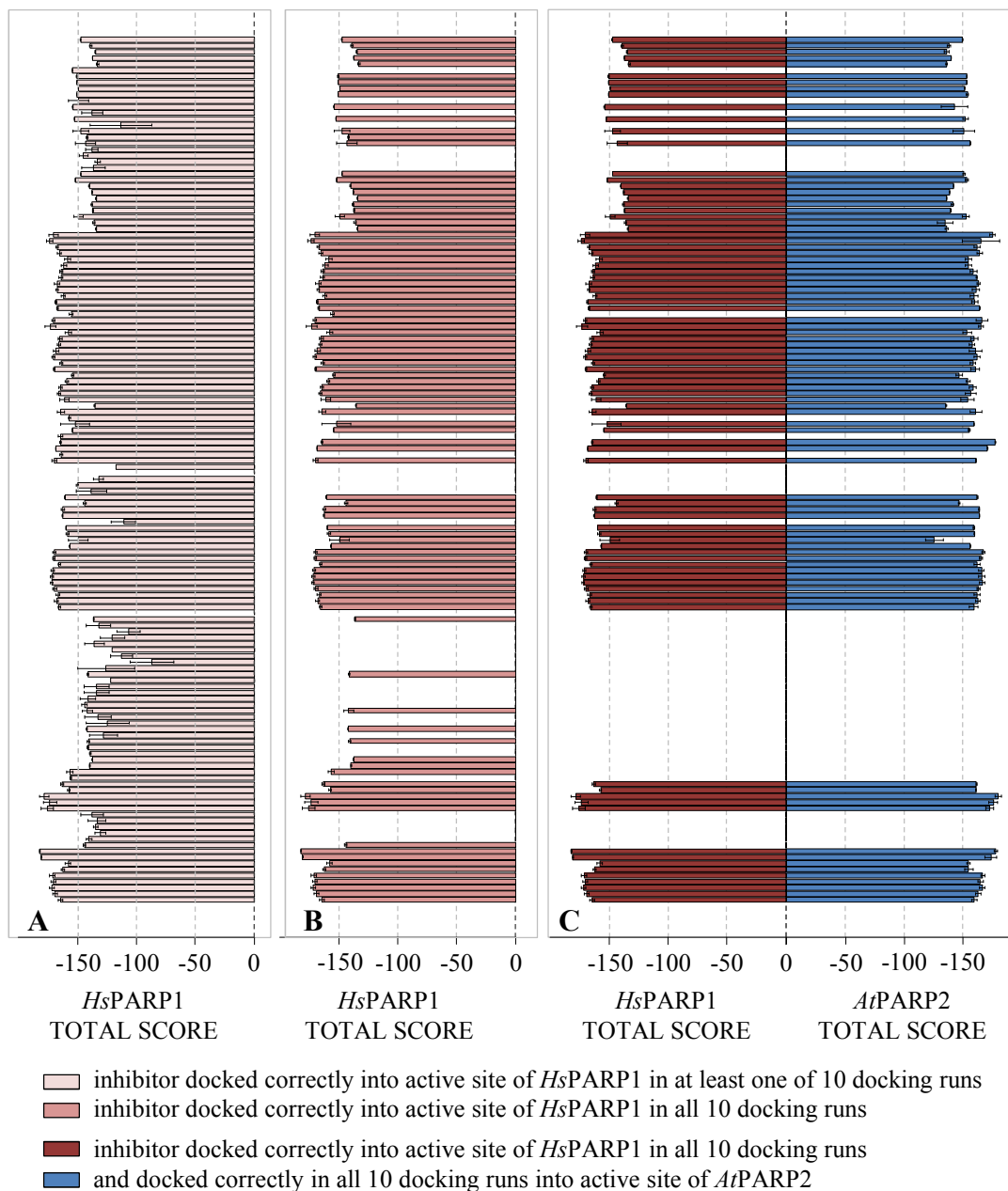
Docking of 142 *Hs*PARP1 inhibitors into *Hs*PARP1 and *At*PARP2

Figure 5.9: *Hs*PARP1 and *At*PARP2 docking scores of 142 *Hs*PARP1 inhibitors

Docking scores represented as bar charts; A and B: docking analysis of inhibitors docked into *Hs*PARP1. C: docking analysis of inhibitors docked into *Hs*PARP1 and *At*PARP2)

5.7 Screening results

5.7.1 Pharmacophore screening results

	KeyOrganics database	KeyOrganics QUIN	KeyOrganics PHTH
Unique structures	43.179	59	41
Unique structures passed	2.713	44	38
Unique structures not passed	40.466	15	3
Tautomers	57.117	136	43
Tautomers passed	2.879	55	40

Substructure search for phthalazinone and quinazolinone cores

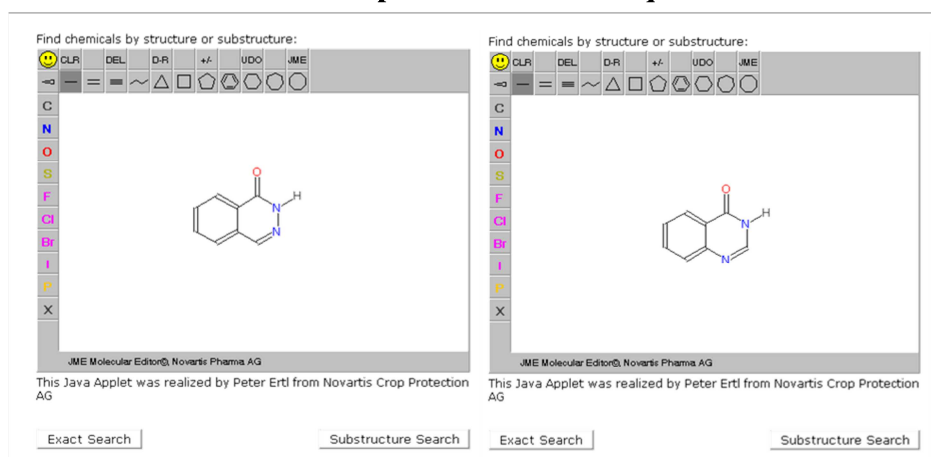


Figure 5.10: Substructure search for phthalazinone and quinazolinone
The search was performed online at shop.keyorganics.co.uk

5.8 ProBiS – conserved amino acids

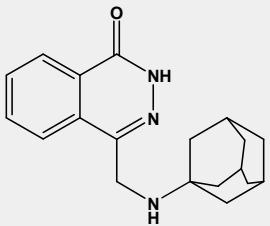
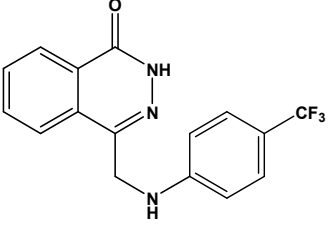
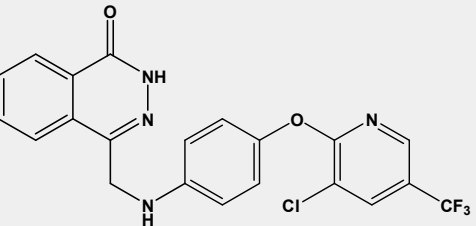
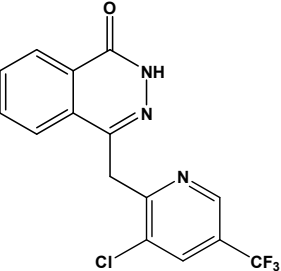
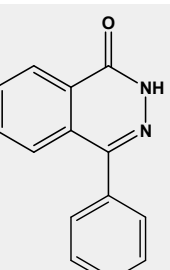
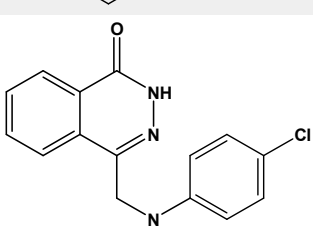
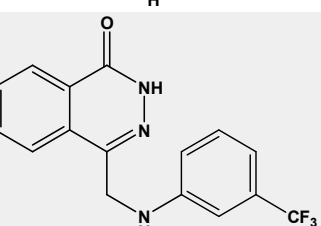
1TOX	1A26	1TOX	1A26	1TOX	1A26	1TOX	1A26
Ile ₁₅₀	Ile ₉₉₀	Lys ₂₄	Arg ₈₆₅	Gln ₃₆	Arg ₈₇₈	Ala ₆₂	Ser ₉₀₄
Tyr ₂₀	Trp ₈₆₁	Ile ₃₁	Ile ₈₇₂	Gly ₅₂	Gly ₈₉₄	Tyr ₆₅	Tyr ₉₀₇
His ₂₁	His ₈₆₂	Gly ₃₄	Gly ₈₇₆	Tyr ₅₄	Tyr ₈₉₆	Glu ₁₄₈	Glu ₉₈₈
Gly ₂₂	Gly ₈₆₃	Ile ₃₅	Leu ₈₇₇	Thr ₅₆	Ala ₈₉₈	Tyr ₁₄₉	Tyr ₉₈₉
Thr ₂₃	Ser ₈₆₄						

5.9 Inhibitors

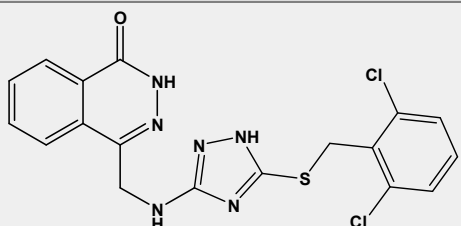
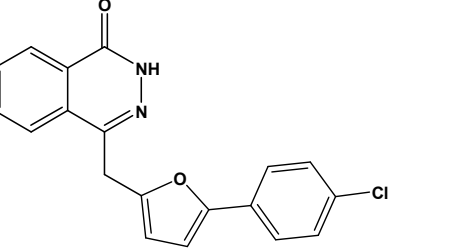
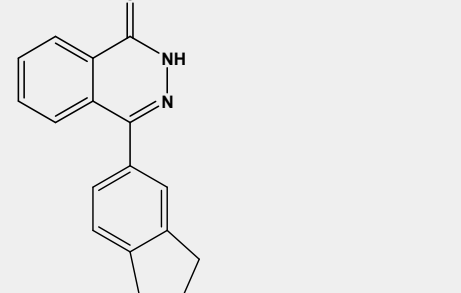
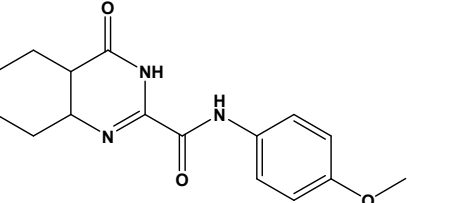
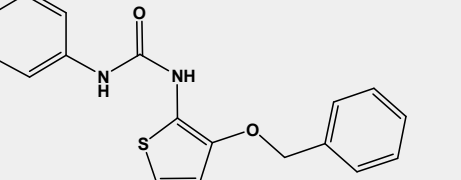
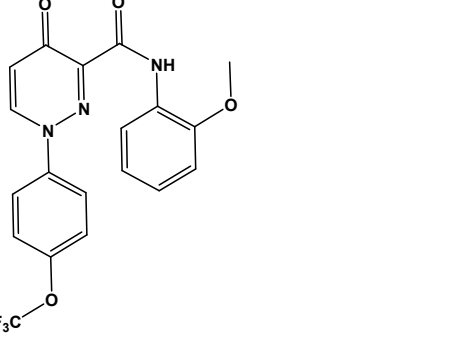
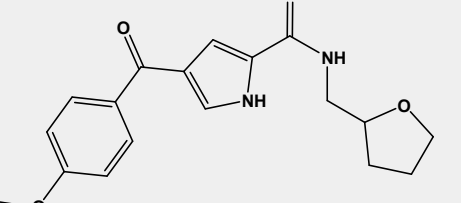
no	structure	IUPAC name	Bionet name
25		4-[(3-{[3-chloro-5-(trifluoromethyl)-2-pyridinyl]oxy}anilino)methyl]-1(2H)-phthalazinone	10E-062
26		N,2-bis[3-(trifluoromethyl)phenyl]-1,3-thiazole-4-carboxamide	12F-408S
27		2-(methylsulfanyl)-4(3H)-quinazolinone	7W-0349
28		2-(2-chlorobenzyl)-4(3H)-quinazolinone	2F-311S
29		2-[(4-benzylpiperidino)methyl]-4(3H)-quinazolinone	MS-3199
30		2-(4-chloroanilino)-4(3H)-quinazolinone	1P-365S
31		2-(4-methylpiperazino)-4(3H)-quinazolinone	1P-333S
32		2-(allylsulfanyl)-4(3H)-quinazolinone	10N-335S
33		2-(4-chlorobenzyl)-4(3H)-quinazolinone	1F-390S

no	structure	IUPAC name	Bionet name
34		2-[(4-benzylpiperazino)methyl]-4(3H)-quinazolinone	MS-3198
35		2-(4-methoxyanilino)-4(3H)-quinazolinone	1P-363S
36		2-(1-pyrrolidinyl)-4(3H)-quinazolinone	1P-330S
37		2-(2-propynylsulfanyl)-4(3H)-quinazolinone	10N-377S
38		2-(3-chlorobenzyl)-4(3H)-quinazolinone	2F-310S
39		2-(morpholino-methyl)-4(3H)-quinazolinone	MS-3180
40		2-(3-methoxyanilino)-4(3H)-quinazolinone	2P-303S
41		2-(1-bromoethyl)-4(3H)-quinazolinone	MS-2995
42		2-[(3,4,4-trifluoro-3-butenyl)sulfanyl]-4(3H)-quinazolinone	7N-764
43		2-[3-(trifluoromethyl)benzyl]-4(3H)-quinazolinone	2F-301S

no	structure	IUPAC name	Bionet name
44		2-(1-morpholino-ethyl)-4(3H)-quinazolinone	MS-2996
45		2-[3-(trifluoromethyl)anilino]-4(3H)-quinazolinone	2P-301S
46		2-(phenoxyethyl)-4(3H)-quinazolinone	MS-3187
47		4-(anilinomethyl)-1(2H)-phthalazinone	12B-072
48		4-[(4-hydroxyanilino)methyl]-1(2H)-phthalazinone	7E-032
49		4-[(4-phenoxyanilino)methyl]-1(2H)-phthalazinone	5D-035
50		4-(2,4-dichlorobenzyl)-1(2H)-phthalazinone	1E-057
51		2-(4-oxo-3,4-dihydro-1-phthalazinyl)acetonitrile	3R-0801

no	structure	IUPAC name	Bionet name
52		4-[(1-adamantylamino)methyl]-1(2H)-phthalazinone	8D-009
53		4-{[4-(trifluoromethyl)anilino]methyl}-1(2H)-phthalazinone	9D-009
54		4-{[4-{[3-chloro-5-(trifluoromethyl)-2-pyridinyl]oxy}anilino]methyl}-1(2H)-phthalazinone	7E-019
55		4-{[3-chloro-5-(trifluoromethyl)-2-pyridinyl]methyl}-1(2H)-phthalazinone	11F-056
56		4-phenyl-1(2H)-phthalazinone	MS-1294
57		4-(4-chloroanilino)methyl-1(2H)-phthalazinone	12B-013
58		4-{[3-(trifluoromethyl)anilino]methyl}-1(2H)-phthalazinone	12B-089

no	structure	IUPAC name	Bionet name
59		4-([5-(2-chlorophenyl)-2-furyl]methyl)-1(2H)-phthalazinone	6F-014
60		4-(5-methyl-2-thienyl)-1(2H)-phthalazinone	9W-0251
61		4-([4-methoxyanilino]methyl)-1(2H)-phthalazinone	12B-093
62		4-([3-(methylsulfanyl)anilino]methyl)-1(2H)-phthalazinone	2C-022
63		4-([5-(3-chlorophenyl)-2-furyl]methyl)-1(2H)-phthalazinone	6F-012
64		4-([4-(trifluoromethoxy)anilino]methyl)-1(2H)-phthalazinone	8D-136
65		N,N-dimethyl-4-([(4-oxo-3,4-dihydro-1-phthalazinyl)methyl]amino)benzenesulfonamide	7D-145

no	structure	IUPAC name	Bionet name
66			8D-022
67		4-([5-(4-chlorophenyl)-2-furyl]methyl)-1(2H)-phthalazinone	6F-011
68		4-(2,3-dihydro-1H-inden-5-yl)-1(2H)-phthalazinone	6W-0242
69		4-methoxy-N-(4-oxo-3,4,5,6,7,8-hexahydro-2-quinazolinyl)benzenecarboxamide	3N-301S
70		N-[3-(benzyloxy)-2-thienyl]-N'-phenylurea	7F-028
71		N-(2-methoxyphenyl)-4-oxo-1-[4-(trifluoromethoxy)phenyl]-1,4-dihydro-3-pyridazinecarboxamide	6P-518S
72		4-(4-methoxybenzoyl)-N-(tetrahydro-2-furanylmethyl)-1H-pyrrole-2-carboxamide	8R-0216

no	structure	IUPAC name	Bionet name
73		4-[[4-(4-chlorophenyl)piperazino]methyl]-1(2H)-phthalazinone	8D-003
74		4-[(4-phenylpiperazino)methyl]-1(2H)-phthalazinone	5D-015
75		4-[(methylanilino)methyl]-1(2H)-phthalazinone	12B-099
76		3-methoxy-N-(4-oxo-3,4,5,6,7,8-hexahydro-2-quinazolinyloxy)benzenecarboxamide	2N-324S
77		4-methyl-N-(4-oxo-3,4,5,6,7,8-hexahydro-2-quinazolinyloxy)benzenecarboxamide	2N-326S
78		4-chloro-N-(4-oxo-3,4,5,6,7,8-hexahydro-2-quinazolinyloxy)benzenecarboxamide	2N-322S
79		2,4-dichloro-N-(4-oxo-3,4,5,6,7,8-hexahydro-2-quinazolinyloxy)benzenecarboxamide	3N-304S

no	structure	IUPAC name	Bionet name
80		2-chloro-N-(4-oxo-3,4,5,6,7,8-hexahydro-2-quinazoliny) benzenecarboxamide	3N-313S
81		3-methyl-N-(4-oxo-3,4,5,6,7,8-hexahydro-2-quinazoliny) benzenecarboxamide	2N-323S
82		3-chloro-N-(4-oxo-3,4,5,6,7,8-hexahydro-2-quinazoliny) benzenecarboxamide	3N-315S
83		2,6-difluoro-N-(4-oxo-3,4,5,6,7,8-hexahydro-2-quinazoliny) benzenecarboxamide	3N-314S
84		4-fluoro-N-(4-oxo-3,4,5,6,7,8-hexahydro-2-quinazoliny) benzenecarboxamide	2N-321S
85		2-methyl-N-(4-oxo-3,4,5,6,7,8-hexahydro-2-quinazoliny) benzenecarboxamide	3N-312S
86		4-oxo-N,1-diphenyl-1,4-dihydro-3-pyridazinecarboxamide	6P-538S
87		1-(4-chlorophenyl)-4-oxo-N-phenyl-1,4-dihydro-3-pyridazinecarboxamide	6P-545S
88		N-(2-methoxyphenyl)-4-oxo-1-phenyl-1,4-dihydro-3-pyridazinecarboxamide	6P-506S

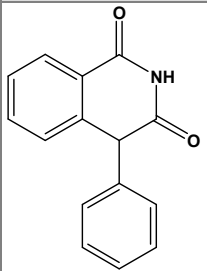
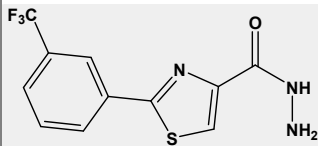
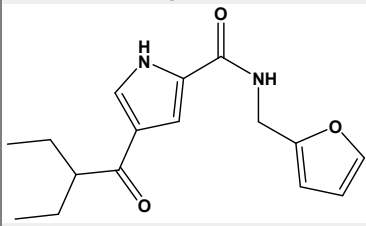
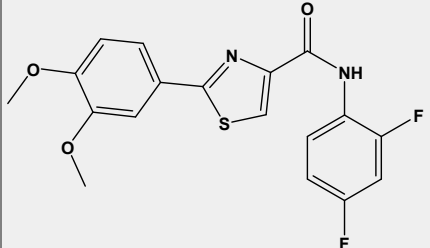
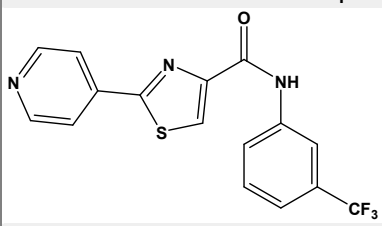
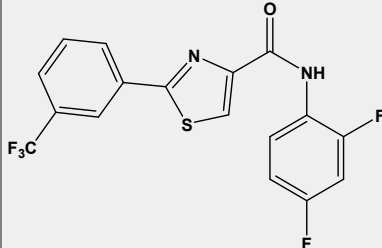
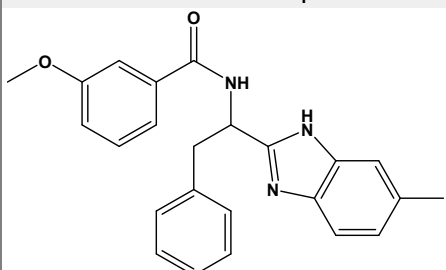
no	structure	IUPAC name	Bionet name
89		1-(4-chlorophenyl)-N-(2-methoxyphenyl)-4-oxo-1,4-dihydro-3-pyridazinecarboxamide	6P-502S
90		1-(4-chlorophenyl)-N-(2-methylphenyl)-4-oxo-1,4-dihydro-3-pyridazinecarboxamide	6P-527S
91		2-[(4-ethylanylino)methyl]-4(3H)-quinazolinone	MS-3206
92		1-(4-chlorophenyl)-4-oxo-1,4-dihydro-3-pyridazinecarbohydrazide	5P-528S
93		1-(4-methoxyphenyl)-4-oxo-1,4-dihydro-3-pyridazinecarboxylic acid	7P-655S
94		4-oxo-1-phenyl-1,4-dihydro-3-pyridazinecarboxylic acid	6P-658S
95		4-oxo-1-[3-(trifluoromethyl)phenyl]-1,4-dihydro-3-pyridazinecarboxylic acid	6P-676S
96		4-hydroxy-6-oxo-1-[3-(trifluoromethyl)phenyl]-1,6-dihydro-3-pyridazinecarboxylic acid	5D-048
97		4-({[(2,4-dichlorobenzyl)oxy]amino}methyl)-1(2H)-phthalazinone	8D-128

no	structure	IUPAC name	Bionet name
98		1-methyl-5-[[[(4-oxo-3,4-dihydro-1-phthalazinyl)methyl]amino]-1H-pyrazole-4-carbonitrile	5D-125
99		4-[(4-nitroanilino)methyl]-1(2H)-phthalazinone	2C-021
100		4-benzyl-1(2H)-phthalazinone	7D-022
101		2-(allylamino)-4(3H)-quinazolinone	1P-303S
102		2-(propylamino)-4(3H)-quinazolinone	1P-345S
103		2-(phenyl)-6-(chloro)-4(3H)-quinazolinone	12J-553
104		2-(3,4-dimethylanilino)-4(3H)-quinazolinone	1P-374S
105		2-(3,5-dimethylanilino)-4(3H)-quinazolinone	2P-302S
106		2-[[[(4-chlorophenyl)sulfonyl]methyl]-4(3H)-quinazolinone	MS-3241

No	structure	IUPAC name	Bionet name
107		2-(trifluoromethyl)-4(3H)-quinazolinone	9L-021
108		ethyl 2-[(4-oxo-3,4-dihydro-2-quinazoliny) sulfanyl]acetate	MS-3625
109		2-[1-(phenethylamino)ethyl]-4(3H)-quinazolinone	MS-3001
110		2-[(2-chlorobenzyl) amino]methyl}-4(3H)-quinazolinone	MS-3205
111		2-[(4-chlorophenyl) sulfanyl]methyl}-4(3H)-quinazolinone	MS-3233
112		2-[(3-pyridinylmethyl) amino]methyl}-4(3H)-quinazolinone	MS-3203
113		2,2,2-trifluoro-N-(7-oxo-4,5,6,7-tetrahydrothieno[2,3-c]pyridin-4-yl)acetamide	10T-0068
114		6-[[2-(4-methoxyphenyl)-2-oxoethyl]sulfanyl]-1-phenyl-1,5-dihydro-4H-pyrazolo[3,4-d]pyrimidin-4-one	9N-663S

No	structure	IUPAC name	Bionet name
115		6-([2-(4-chlorophenyl)-2-oxoethyl]sulfanyl)-1-phenyl-1,5-dihydro-4H-pyrazolo[3,4-d]pyrimidin-4-one	9N-661S
116		6-([2-(2-oxo-2-phenylethyl)sulfanyl]-1-phenyl-1,5-dihydro-4H-pyrazolo[3,4-d]pyrimidin-4-one	9N-665S
117		3-(1H-1,2,3,4-tetraazol-5-yl)pyrazolo[1,5-a]quinazolin-5(4H)-one	6W-0853
118		6-(decyloxy)nicotinic acid	12N-180
119		2-(3,4,4-trifluorobut-3-enylthio)benzoic acid	6N-761
120		2-[(3,4,4-trifluoro-3-butenyl)sulfanyl]nicotinic acid	7N-744
121		2-[(4-chlorobenzyl)sulfanyl]benzenecarboxylic acid	8K-511S
122		3-{[3-(trifluoromethyl)benzyl]sulfanyl}-2-thiophenecarboxylic acid	10G-320

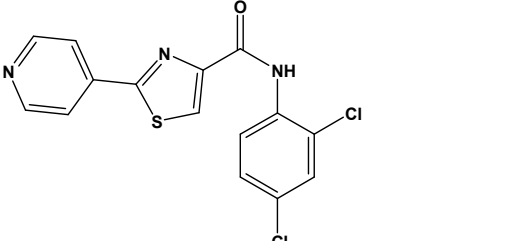
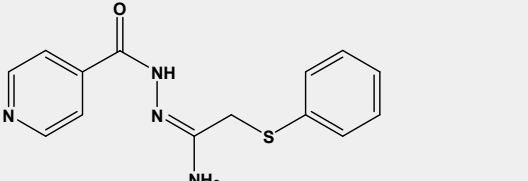
No	structure	IUPAC name	Bionet name
123		3-[[2-(benzoylamino)acetyl]amino]-2-thiophenecarboxylic acid	7F-938
124		4-(2,3-dichlorobenzoyl)-1-methyl-1H-pyrrole-2-carboxylic acid	9R-0349
125		3-{4-[3-(trifluoromethyl)phenyl]-1H-pyrazol-1-yl}benzenecarboxylic acid	8J-583S
126		2-[[5-chloro-1-methyl-3-phenyl-1H-pyrazol-4-yl)methyl]sulfanyl}benzenecarboxylic acid	3K-634S
127		2-(4-fluorophenyl)imidazo[1,2-a]pyridine-8-carboxylic acid	6X-0839
128		2-({2-[(4-chlorophenyl)sulfanyl]acetyl}amino)benzenecarboxamide	MS-3232
129		3-[(4-{[3-chloro-5-(trifluoromethyl)-2-pyridinyl]methyl}phenoxy)methyl]benzenecarboxamide	7N-774

No	structure	IUPAC name	Bionet name
130		4-phenyl-1,3(2H,4H)-isoquinolinedione	9G-036
131		2-(3-(trifluoromethyl)phenyl)thiazole-4-carbohydrazide	5G-439S
132		4-(2-ethylbutanoyl)-N-(2-furylmethyl)-1H-pyrrole-2-carboxamide	8R-0345
133		N-(2,4-difluorophenyl)-2-(3,4-dimethoxyphenyl)-1,3-thiazole-4-carboxamide	3G-337S
134		2-(4-pyridinyl)-N-[3-(trifluoromethyl)phenyl]-1,3-thiazole-4-carboxamide	1G-351S
135		N-(2,4-difluorophenyl)-2-[3-(trifluoromethyl)phenyl]-1,3-thiazole-4-carboxamide	12F-409S
136		3-methoxy-N-[1-(5-methyl-1H-1,3-benzimidazol-2-yl)-2-phenylethyl]benzenecarboxamide	3J-311S

No	structure	IUPAC name	Bionet name
137		N-[1-(1H-1,3-benzimidazol-2-yl)-2-phenylethyl]benzenecarboxamide	2J-378S
138		N-[1-(1H-1,3-benzimidazol-2-yl)-2-phenylethyl]-2-fluorobenzenecarboxamide	3J-341S
139		N-benzoyl-N'-(2-phenoxyphenyl)thiourea	9N-068S
140		N-[3-(benzyloxy)-2-thienyl]-2-[3-chloro-5-(trifluoromethyl)-2-pyridinyl]-1-hydrazine carboxamide	11F-021
141		2-[3-chloro-5-(trifluoromethyl)-2-pyridinyl]-N-{3-[(2,4-dichlorobenzyl)oxy]-2-thienyl}-1-hydrazine carboxamide	11F-023
142		N'-[3-chloro-5-(trifluoromethyl)-2-pyridinyl]-4-{2-[3-chloro-5-(trifluoromethyl)-2-pyridinyl]propanoyl}-1H-pyrrole-2-carbohydrazide	8J-019
143		3-{1-[(2-fluoro[1,1'-biphenyl]-4-yl)oxy]ethyl}-N-(4-methylphenyl)-1H-pyrazole-1-carboxamide	12P-109

No	structure	IUPAC name	Bionet name
144		3-[(2,4-dichlorobenzyl)sulfanyl]-N-(4-fluorophenyl)-2-thiophenecarboxamide	10G-507S
145		N'-[3-chloro-5-(trifluoromethyl)-2-pyridinyl]-4-(2,4-dichlorobenzoyl)-1H-pyrrole-2-carbohydrazide	3G-035
146		4-(2-methylbenzoyl)-N-[2-(2-pyridinyl)ethyl]-1H-pyrrole-2-carboxamide	9R-0237
147		2-[(4-chlorophenyl)sulfanyl]-N'-isonicotinoyl ethanehydrazonamide	2F-035
148		N-allyl-4-{2-[3-chloro-5-(trifluoromethyl)-2-pyridinyl]propanoyl}-1H-pyrrole-2-carboxamide	8J-020
149		4-(2,3-dichlorobenzoyl)-N-[3-(dimethylamino)propyl]-1H-pyrrole-2-carboxamide	1R-0015
150		N-(2-methoxybenzyl)-4-(2-methylbenzoyl)-1H-pyrrole-2-carboxamide	9R-0277
151		N-(2-furylmethyl)-4-(3-methoxybenzoyl)-1H-pyrrole-2-carboxamide	7R-0329

No	structure	IUPAC name	Bionet name
152		1-(6-methyl-2-pyridinyl)-N'-[(E)-phenylmethylidene]-1H-imidazole-4-carbohydrazide	8P-705
153		N'-[(E)-(3,4-dimethoxyphenyl)methylidene]-3-[(4-fluorobenzyl)oxy]-2-thiophenecarbohydrazide	9N-031
154		3-(1H-1,2,4-triazol-1-ylmethyl)-N-[4-(trifluoromethoxy)phenyl]-1H-1,2,4-triazole-5-carboxamide	3F-040
155		N-{5-[(2-methylbenzyl)sulfanyl]-1H-1,2,4-triazol-3-yl}-2-thiophenecarboxamide	MS-3260
156		N'-[3-chloro-5-(trifluoromethyl)-2-pyridinyl]-4-(2,4-dichlorobenzoyl)-1H-pyrrole-2-carbohydrazide	3D-035
157		4-chloro-N'-[(4-oxo-3,4-dihydro-1-phthalazinyl)methyl]benzene carbohydrazide	2C-005
158		2-chloro-N-[1-(4-oxo-3,4-dihydro-2-quinazoliny)ethyl]-N-pentylacetamide	MS-3033
159		2-(3-pyridinyl)-N-[3-(trifluoromethyl)phenyl]-1,3-thiazole-4-carboxamide	2G-327S

No	structure	IUPAC name	Bionet name
160		N-(3,5-dichlorophenyl)-2-(4-pyridinyl)-1,3-thiazole-4-carboxamide	1G-395S
161		N'-isonicotinoyl-2-(phenylsulfanyl)ethanehydrazonamide	2F-072

5.10 Inhibitors; *in silico* and *in vitro* results

5.10.1 Docking results, *in silico* and *in vitro* results

No	Docking score	Inhibition 100 μ M (%)	IC ₅₀ \pm SE (μ M)	theoret.	<i>in vitro</i>	result
25	-163.2770	62.8	>100	positive	positive	TP
26	-102.2740	10.7	>100	negative	negative	TN
27	-136.2320	75.0	1.46 \pm 0.152	negative	positive	FN
28	-151.5510	1.1	79.4 \pm 1.05	positive	negative	FP
29	-155.1420	34.3	57.9 \pm 1.10	positive	negative	FP
30	Not available					
31	Not available					
32	-143.0310	73.5	3.62 \pm 1.84	negative	positive	FN
33	-148.1600	20.2	>100	positive	negative	FP
34	-153.6020	81.4	6.28 \pm 1.18	positive	positive	TP
35	-98.8115	84.2	1.79 \pm 0.312	negative	positive	FN
36	-138.6550	71.1	11.1 \pm 1.22	negative	positive	FN
37	-142.6430	30.3	15.5 \pm 1.28	negative	negative	FP
38	Not available					
39	-149.4910	61.6	19.8 \pm 1.12	positive	positive	TP
40	--132.2260	42.9	18.3 \pm 1.17	negative	positive	FN
41	-139.6590	80.2	10.9 \pm 1.14	negative	positive	FN
42	-145.3650	98.0	19.66 \pm 1.22	positive	positive	TP
43	-158.9430	0.8	>100	positive	negative	FP
44	-141.5430	48.0	55.2 \pm 1.23	negative	positive	FN
45	Not available					
46	-153.1800	42.9	31.09 \pm 0.09	positive	positive	TP
47	-154.1540	62.8	0.08 \pm 0.46	positive	positive	TP
48	-156.0940	44.8	3.62 \pm 1.80	positive	positive	TP
49	-155.7730	106.3	56.1 \pm 1.08	positive	positive	TP
50	-158.5070	36.5	>100	positive	negative	FP
51	-127.3730	68.1	0.707 \pm 0.21	negative	positive	FN
52	-160.0420	199.9	1.62 \pm 1.17	positive	positive	TP
53	-156.7790	151.3	1.13 \pm 0.28	positive	positive	TP
54	-169.5060	149.8	6.25 \pm 1.24	positive	positive	TP
55	-164.1940	40.6	8.31 \pm 1.19	positive	positive	TP
56	-146.2120	73.5	0.529 \pm 0.126	positive	positive	TP
57	Not available					
58	-163.6030	56.7	1.97 \pm 0.21	positive	positive	TP
59	-159.4200	13.4	18.4 \pm 1.27	positive	negative	FP
60	-147.0350	174.0	6.38 \pm 1.09	positive	positive	TP

No	Docking score	Inhibition 100 μ M (%)	IC ₅₀ \pm SE (μ M)	theoret.	<i>in vitro</i>	result
61	-155.2210	44.8	1.01 \pm 0.113	positive	positive	TP
62	-158.3430	135.2	13.4 \pm 1.22	positive	positive	TP
63	-157.6220	39.7	>100	positive	negative	FP
64	-157.8790	134.6	4.45 \pm 1.35	positive	positive	TP
65	-148.7220	149.8	1.79 \pm 0.312	positive	positive	TP
66	Not available					
67	-159.0210	35.2	>100	positive	negative	FP
68	-152.6610	24.2	0.81 \pm 0.07	positive	negative	FP
69	-139.7300	<0	>100	negative	negative	TN
70	Not available					
71	-131.5060	29.8	>100	negative	negative	TN
72	-140.0460	<0	>100	negative	negative	TN
73	-160.5560	138.6	59.0 \pm 1.13	positive	positive	TP
74	Not available					
75	Not available					
76	-149.0080	9.2	20.1 \pm 1.13	positive	negative	FP
77	-143.7000	82.7	58.5 \pm 1.11	negative	positive	FN
78	-140.2210	<0	>100	negative	negative	TN
79	-139.9130	<0	>100	negative	negative	TN
80	-116.4390	<0	>100	negative	negative	TN
81	-144.4020	68.0	11.05 \pm 1.28	negative	positive	FN
82	-145.5410	44.3	>100	positive	positive	TP
83	-135.4130	75.8	20.7 \pm 1.08	negative	positive	FN
84	-145.1870	66.9	15.5 \pm 1.28	positive	positive	TP
85	-140.3310	50.8	14.3 \pm 1.69	negative	positive	FN
86	-124.3890	3.1	>100	negative	negative	TN
87	-127.3480	1.5	>100	negative	negative	TN
88	-129.1940	36.8	>100	negative	negative	TN
89	-84.8195	17.3	>100	negative	negative	TN
90	-130.5720	3.9	>100	negative	negative	TN
91	Not available					
92	-128.5490	<0	>100	negative	negative	TN
93	-143.3980	2.7	>100	negative	negative	TN
94	-116.1240	0.6	>100	negative	negative	TN
95	-142.8430	0.6	>100	negative	negative	TN
96	-169.4140	24.8	>100	positive	negative	FP
97	-161.9990	34.3	18.93 \pm 5.27	positive	negative	FP
98	-154.5820	49.9	>100	positive	positive	TP
99	-153.0150	148.1	11.2 \pm 1.41	positive	positive	TP
100	-152.3980	58.8	0.522 \pm 0.33	positive	positive	TP

No	Docking score	Inhibition 100 μ M (%)	IC ₅₀ \pm SE (μ M)	theoret.	<i>in vitro</i>	result
101	-140.7690	75.3	1.76 \pm 1.49	negative	positive	FN
102	-140.7500	73.5	58.1 \pm 1.13	negative	positive	FN
103	-134.6420	<0	>100	negative	negative	TN
104	-133.5520	2.8	76.1 \pm 1.04	negative	negative	TN
105	-132.5300	22.2	>100	negative	negative	TN
106	-146.9410	104.9	2.77 \pm 1.46	positive	positive	TP
107	-141.6250	14.3	>100	negative	negative	TN
108	-151.5050	76.8	0.93 \pm 0.06	positive	positive	TP
109	-161.3500	83.5	3.74 \pm 1.17	positive	positive	TP
110	-161.0260	55.4	4.62 \pm 1.18	positive	positive	TP
111	-159.5130	5.8	>100	positive	negative	FP
112	-157.6700	86.9	1.58 \pm 1.20	positive	positive	TP
113	-148.2510	90.4	13.6 \pm 1.55	positive	positive	TP
114	-150.6640	0.2	>100	positive	negative	FP
115	-158.5720	3.5	>100	positive	negative	FP
116	-149.6210	56.6	>100	positive	positive	TP
117	-146.3320	12.8	9.68 \pm 1.13	positive	negative	FP
118	-150.4170	3.5	>100	positive	negative	FP
119	-146.4570	14.2	>100	positive	negative	FP
120	-128.1530	7.8	>100	negative	negative	TN
121	-152.4840	<0	>100	positive	negative	FP
122	-157.8750	2.5	>100	positive	negative	FP
123	-125.5620	0.4	>100	negative	negative	TN
124	-123.2550	<0	>100	negative	negative	TN
125	-132.6790	0.7	>100	negative	negative	TN
126	-151.7160	2.6	>100	positive	negative	FP
127	-156.9440	42.0	>100	positive	positive	TP
128	-156.2550	13.9	>100	positive	negative	FP
129	-171.9060	2.3	58.5 \pm 7.34	positive	negative	FP
130	-153.9730	<0	>100	positive	negative	FP
131	-151.7840	<0	>100	positive	negative	FP
132	-150.3260	<0	>100	positive	negative	FP
133	-127.6170	1.2	>100	negative	negative	TN
134	-128.2610	<0	>100	negative	negative	TN
135	-101.3870	<0	>100	negative	negative	TN
136	-145.0710	0.8	>100	positive	negative	FP
137	-109.3430	1.7	>100	negative	negative	TN
138	-116.5060	2.7	>100	negative	negative	TN
139	-125.3950	0.4	>100	negative	negative	TN
140	-105.9610	12.0	>100	negative	negative	TN

No	Docking score	Inhibition 100 μ M (%)	IC ₅₀ \pm SE (μ M)	theoret.	<i>in vitro</i>	result
141	-93.5362	7.0	8.31 \pm 1.19	negative	negative	TN
142	-114.5590	<0	>100	negative	negative	TN
143	-102.7840	35.9	>100	negative	negative	TN
144	-101.6340	8.8	>100	negative	negative	TN
145	Not available					
146	-145.7420	<0	>100	positive	negative	FP
147	-145.5470	<0	>100	positive	negative	FP
148	-146.1270	<0	>100	positive	negative	FP
149	-144.4850	21.6	>100	negative	negative	TN
150	-134.7230	<0	>100	negative	negative	TN
151	-136.1820	21.9	>100	negative	negative	TN
152	-85.9073	<0	>100	negative	negative	TN
153	-123.8610	14.2	>100	negative	negative	TN
154	-133.3860	<0	>100	negative	negative	TN
155	-128.8580	3.4	>100	negative	negative	TN
156	-137.4650	10.0	>100	negative	negative	TN
157	Not available					
158	Not available					
159	Not available					
160	Not available					
161	Not available					

5.10.2 POSIT results

No.	Rec.	Result	Prob.	M166	TC	LS	No.	Rec.	Result	Prob.	M166	TC	LS
100	3C49	*	0.50	0.59	1.29	0.00	65	3C49	*	0.42	0.51	1.19	7.84
51	3C4H	**	0.72	0.59	1.42	1.66	54	1UK1	*	0.50	0.59	1.14	8.84
27	3C4H	***	0.93	0.62	1.77	1.16	73	3C49	***	0.85	0.76	1.31	8.21
32	3C4H	***	0.83	0.62	1.66	4.52	52	3GJW	**	0.51	0.67	1.08	5.40
37	3C4H	***	0.83	0.64	1.65	1.10	97	3GJW	**	0.72	0.55	1.36	5.26
36	3C4H	***	0.90	0.65	1.62	0.00	64	3GJW	*	0.50	0.56	1.16	7.77
56	3C4H	*	0.42	0.50	1.26	0.00	53	3GJW	*	0.50	0.58	1.24	2.04
46	3C4H	**	0.72	0.62	1.35	0.06	59	3C49	*	0.72	0.60	1.34	7.90
47	3GJW	*	0.50	0.62	1.21	7.78	55	3GJW	*	0.42	0.48	1.18	1.75
48	3GJW	*	0.50	0.58	1.22	1.84	76	3GJW	*	0.50	0.55	1.22	4.81
84	3GJW	*	0.42	0.53	1.29	5.03	83	3GJW	*	0.42	0.52	1.25	3.48
81	3GJW	*	0.50	0.56	1.29	4.57	129	3L3M	*	0.29	0.54	1.06	5.55
77	3GJW	*	0.50	0.56	1.29	5.22	42	3C4H	***	0.83	0.60	1.64	2.04
85	3GJW	*	0.50	0.55	1.29	3.39	68	3L3L	*	0.42	0.50	1.16	0.01
113	3C4H	*	0.43	0.45	1.14	4.38	117	3GJW	*	0.25	0.45	1.08	0.00
60	3C4H	*	0.42	0.53	1.27	0.42	28	3C4H	**	0.72	0.62	1.47	8.97
101	3C4H	***	0.81	0.55	1.54	9.87	104	3C4H	**	0.75	0.55	1.32	0.00
102	3C4H	***	0.83	0.57	1.55	5.33	44	3C4H	**	0.72	0.67	1.29	7.03
35	3C4H	*	0.42	0.52	1.23	0.01	109	3GJW	*	0.50	0.61	1.30	8.51
40	3C4H	*	0.42	0.52	1.23	5.56	39	3C4H	**	0.72	0.64	1.31	4.34
41	3C4H	***	0.90	0.65	1.62	7.81	34	1UK0	***	0.78	0.76	1.25	7.45
58	3C49	**	0.72	0.66	1.26	3.32	29	1UK0	***	0.78	0.76	1.30	8.41
61	3GJW	*	0.50	0.59	1.24	1.92	112	3GJW	***	0.79	0.69	1.33	2.97
99	3GJW	*	0.42	0.54	1.20	1.78	110	3GJW	**	0.72	0.64	1.30	8.05
62	3C49	*	0.50	0.58	1.21	4.88	106	3GJW	*	0.43	0.39	1.20	5.89
49	3C49	**	0.72	0.64	1.32	7.58	108	3C4H	***	0.83	0.62	1.52	3.76

Abbreviations: Rec.: Receptor file used (PDB entry used as template for homology modelling AtPARP1); Result: (quality) classes: *: poor; ** good; *** great; Prob.: Probability of pose being a bioactive pose; M166: MACCS 166 value; TC: Tanimoto combo value; LS: local strain

5.10.3 Structures used for binary QSAR – training set - actives

No	Inhib. 100 μM (%)	Q_VSA _HYD	Q_VSA _NEG	Q_VSA _PNEG	Q_VSA _POL	Q_VSA _POS	Q_VSA _PPOS	Vdw _area	Vdw _vol
60	174	151.203	129.205	25.811	67.141	89.139	41.329	218.344	308.767
53	151	192.339	164.669	44.298	83.531	111.200	39.233	275.869	379.484
54	150	326.695	189.630	19.788	50.402	187.467	30.614	377.097	512.000
99	148	156.613	167.366	82.720	116.935	106.182	34.216	273.548	369.530
73	139	289.862	236.082	29.697	51.266	105.045	21.569	341.128	464.923
62	135	213.679	151.643	42.488	72.675	134.711	30.188	286.354	392.440
49	106	291.764	190.513	17.579	39.148	140.398	21.569	330.912	471.263
106	105	194.893	158.778	66.265	94.346	130.462	28.081	289.240	385.199
42	98	180.181	147.907	24.381	65.710	97.985	41.329	245.892	314.375
112	87	226.882	163.029	13.841	35.409	99.262	21.569	262.291	361.794
35	84	194.022	147.272	22.027	67.896	114.646	45.869	261.918	357.102
109	83	244.454	174.876	32.588	54.157	123.735	21.569	298.611	419.728
34	81	301.328	213.252	19.386	40.955	129.032	21.569	342.283	474.209
41	80	152.549	89.362	32.451	54.020	117.207	21.569	206.569	268.596
108	77	132.965	144.228	64.882	120.920	109.656	56.038	253.885	322.690
101	75	131.810	118.500	31.778	77.647	90.957	45.869	209.457	274.423
27	75	100.952	104.505	40.110	81.439	77.886	41.329	182.391	237.562
56	74	171.429	141.062	17.442	39.011	69.378	21.569	210.440	304.912
102	74	126.741	111.393	38.271	84.140	99.488	45.869	210.881	280.325
36	71	152.900	114.402	19.386	56.637	95.134	37.250	209.536	292.474
51	68	141.893	124.322	17.442	39.011	56.582	21.569	180.904	239.797
84	67	164.847	133.241	35.299	94.117	125.724	58.819	258.964	362.565
25	63	326.695	212.244	19.788	50.402	164.853	30.614	377.097	512.000
47	63	185.388	138.990	29.834	60.022	106.420	30.188	245.410	345.628

5.10.4 Structures used for binary QSAR – training set – inactives

No	Inhib. 100 μM (%)	Q_VSA _HYD	Q_VSA _NEG	Q_VSA _PNEG	Q_VSA _POL	Q_VSA _POS	Q_VSA _PPOS	Vdw _area	Vdw _vol
89	0.17	263.117	193.031	46.225	76.252	146.338	30.027	339.369	442.426
107	0.14	127.588	106.340	13.704	44.318	65.566	30.614	171.906	224.606
119	0.14	158.817	140.873	32.012	76.805	94.750	44.793	235.622	286.625
128	0.14	183.703	183.133	48.784	120.301	120.870	71.517	304.003	388.786
59	0.13	258.704	185.958	17.442	39.011	111.756	21.569	297.715	427.976
117	0.13	161.486	132.923	17.579	52.032	80.595	34.453	213.518	295.543
26	0.11	229.321	225.109	36.610	100.296	104.508	63.686	329.617	438.775
156	0.10	292.449	181.335	35.426	91.737	202.851	56.311	384.185	487.775

No	Inhib. 100 μM (%)	Q_VSA HYD	Q_VSA NEG	Q_VSA PNEG	Q_VSA POL	Q_VSA POS	Q_VSA PPOS	Vdw area	Vdw vol
76	0.09	190.170	148.834	37.802	96.621	137.958	58.819	286.791	394.513
120	0.08	156.961	143.282	32.012	76.805	90.484	44.793	233.766	277.549
141	0.07	303.310	226.688	42.694	114.798	191.420	72.104	418.108	519.640
111	0.06	200.576	168.598	28.566	69.895	101.873	41.329	270.471	368.492
115	0.04	250.747	202.195	45.839	99.892	148.444	54.053	350.640	474.132
118	0.04	228.881	159.549	62.732	87.765	157.097	25.033	316.646	408.878
155	0.03	195.706	171.637	42.797	112.506	136.575	69.709	308.212	405.748
104	0.03	198.754	149.429	19.523	65.392	114.718	45.869	264.147	370.869
93	0.03	160.430	147.123	49.660	83.151	96.459	33.491	243.582	303.731
138	0.03	295.026	217.263	13.841	44.028	121.791	30.188	339.054	488.986
122	0.03	135.541	156.030	54.612	128.211	107.723	73.599	263.752	338.068
137	0.02	290.617	228.800	13.841	44.028	105.845	30.188	334.645	485.843
87	0.02	237.581	187,206	39.526	69.553	119.927	30.027	307.133	407.335
28	0.01	205.803	164.646	19.386	40.955	82.113	21.569	246.758	346.109
43	0.01	215.319	168.478	13.704	44.318	91.158	30.614	259.636	364.009
136	0.01	337.581	241.557	16.344	46.532	142.556	30.188	384.113	545.363
95	0.01	152.112	157.556	47.156	89.693	84.250	42.537	241.805	302.497
94	0.01	130.698	131.876	47.156	80.647	79.469	33.491	211.346	268.641
123	0.00	134.018	156.088	67.158	155.088	133.018	87.930	289.106	365.620
114	0.00	256.326	225.177	54.915	108.968	140.117	54.053	365.294	493.267
148	0.00	227.945	195.758	72.737	120.428	152.615	47.692	348.373	441.658
92	-0.01	175.907	130.639	43.264	73.291	118.560	30.027	249.198	306.221
132	-0.01	197.186	174.191	66.975	105.621	128.615	38.646	302.807	398.957
150	-0.01	282.675	216.230	31.983	70.629	137.074	38.646	353.304	483.600
134	-0.01	206.052	199.631	36.610	91.250	97.671	54.640	297.302	395.844
154	-0.01	237.626	199.891	16.208	70.221	107.957	54.014	307.847	380.010
124	-0.02	200.239	144.062	37.110	70.601	126.778	33.491	270.840	334.869
103	-0.02	194.254	155.944	13.704	35.272	73.582	21.569	229.526	321.680
152	-0.02	251.330	209.231	36.397	66.496	108.595	30.099	317.826	415.006
135	-0.04	220.991	198.737	36.610	86.985	109.239	50.375	307.976	411.204
131	-0.07	150.499	152.283	40.349	90.724	88.940	50.375	241.223	303.805
78	-0.29	178.019	125.852	35.299	94.117	146.285	58.819	272.137	375.378
142	-0.31	309.279	229.207	56.382	121.739	201.810	65.356	431.017	545.457
121	-0.37	183.396	172.754	32.012	76.805	87.447	44.793	260.202	340.742
80	-0.65	178.019	155.434	35.299	94.117	116.703	58.819	272.137	375.378
130	-1.05	159.574	144.190	27,271	61.789	77.172	34.518	221.362	321.880

5.10.5 Binary QSAR – model selection

A: total accuracy; B: accuracy on actives; C: accuracy on inactives; bold: best value per PC;
red bold: accuracies for selected model with smoothing factor = 0.10 and 5 PCs

A PCs	Smoothing factor				
	0.05	0.10	0.15	0.20	0.25
1	73.52	70.59	69.12	69.12	69.12
2	76.47	72.06	72.06	70.59	69.12
3	82.35	79.41	77.94	79.41	79.41
4	89.71	85.29	82.35	82.35	82.35
5	94.12	92.65	88.23	83.82	86.76
6	94.12	92.65	88.23	83.82	86.76

B PCs	Smoothing factor				
	0.05	0.10	0.15	0.20	0.25
1	29.17	25.00	20.83	20.83	20.83
2	45.83	37.50	33.33	29.17	25.00
3	62.50	54.17	50.00	54.17	54.17
4	75.00	62.50	54.17	54.17	54.17
5	83.33	79.17	70.83	62.50	66.67
6	83.33	79.17	70.83	62.50	66.67

C PCs	Smoothing factor				
	0.05	0.10	0.15	0.20	0.25
1	97.72	95.45	95.45	95.45	95.45
2	93.18	90.91	93.18	93.18	93.18
3	93.18	93.18	93.18	93.18	93.18
4	97.73	97.73	97.73	97.73	97.73
5	100.0	100.0	97.73	95.45	97.72
6	100.0	100.0	97.73	95.45	97.72

5.10.6 Binary QSAR – training set– results

No	prediction		LOO cross-validation		No	prediction		LOO cross-validation	
	PRED	RES	PRED	RES		PRED	RES	PRED	RES
active structures (inhibiting <i>At</i> PARP1 >40% at 100 μ M concentration)									
60 ⁺	0.459	1	0.311	1	34 ⁺	0.538	0	0.350	1
53 ⁺	0.669	0	0.544	0	41 ⁺	0.968	0	0.935	0
54 ⁺	0.917	0	0.831	0	108 ⁺	0.511	0	0.261	1
99 ⁺	0.969	0	0.782	0	101 ⁺	0.763	0	0.621	0
73 ⁺	0.273	1	0.077	1	27 ⁺	0.756	0	0.544	0
62 ⁺	0.621	0	0.393	1	56 ⁺	0.346	1	0.219	1
49 ⁺	0.828	0	0.728	0	102 ⁺	0.865	0	0.763	0
106 ⁺	0.691	0	0.433	1	36 ⁺	0.877	0	0.794	0
42 ⁺	0.247	1	0.100	1	51 ⁺	0.722	0	0.454	1
112 ⁺	0.821	0	0.727	0	84 ⁺	0.528	0	0.174	1
35 ⁺	0.305	1	0.140	1	25 ⁺	0.828	0	0.663	0
109 ⁺	0.789	0	0.576	0	47 ⁺	0.864	0	0.798	0
inactive structures (inhibiting <i>At</i> PARP1 <40% at 100 μ M concentration)									
89 ⁻	0.005	0	0.009	0	43 ⁻	0.097	0	0.125	0
107 ⁻	0.455	0	0.735	-1	136 ⁻	0.003	0	0.006	0
119 ⁻	0.008	0	0.015	0	95 ⁻	0.035	0	0.053	0
128 ⁻	0.025	0	0.038	0	94 ⁻	0.247	0	0.380	0
59 ⁻	0.197	0	0.253	0	123 ⁻	0.001	0	0.004	0
117 ⁻	0.309	0	0.430	0	114 ⁻	0.008	0	0.014	0
26 ⁻	0.002	0	0.004	0	148 ⁻	0.001	0	0.002	0
156 ⁻	0.006	0	0.020	0	92 ⁻	0.016	0	0.026	0
76 ⁻	0.258	0	0.378	0	132 ⁻	0.000	0	0.001	0
120 ⁻	0.009	0	0.015	0	150 ⁻	0.120	0	0.156	0
141 ⁻	0.000	0	0.001	0	134 ⁻	0.125	0	0.233	0
111 ⁻	0.080	0	0.101	0	154 ⁻	0.001	0	0.004	0
115 ⁻	0.102	0	0.141	0	124 ⁻	0.026	0	0.048	0
118 ⁻	0.003	0	0.011	0	103 ⁻	0.100	0	0.153	0
155 ⁻	0.010	0	0.014	0	152 ⁻	0.003	0	0.005	0
104 ⁻	0.317	0	0.426	0	135 ⁻	0.012	0	0.017	0
93 ⁻	0.004	0	0.005	0	131 ⁻	0.012	0	0.018	0
138 ⁻	0.024	0	0.035	0	78 ⁻	0.004	0	0.009	0
122 ⁻	0.020	0	0.044	0	142 ⁻	0.023	0	0.055	0
137 ⁻	0.047	0	0.082	0	121 ⁻	0.032	0	0.047	0
87 ⁻	0.100	0	0.137	0	80 ⁻	0.132	0	0.178	0
28 ⁻	0.160	0	0.219	0	130 ⁻	0.067	0	0.124	0

-: inactive compound; +: active compound

5.10.7 Binary QSAR – external validation – results

No	IC ₅₀ (μM)	Inhib. 100 μM (%)	active	PRED	Pred. activity	decision
52	1.62	199.9	Yes	0.014	Inactive	Wrong
65	1.79	149.8	Yes	0.039	Inactive	Wrong
64	4.45	134.6	Yes	0.332	Inactive	Wrong
113	13.6	90.4	Yes	0.399	Inactive	Wrong
77	58.5	82.7	Yes	0.639	Active	Correct
83	20.7	75.8	Yes	0.214	Inactive	Wrong
32	3.62	73.5	Yes	0.492	Inactive	Wrong
81	11.05	68.0	Yes	0.639	Active	Correct
39	19.8	61.6	Yes	0.797	Active	Correct
153	>100	14.2	No	0.046	Inactive	Correct
140	>100	12.0	No	0.003	Inactive	Correct
144	>100	8.8	No	0.007	Inactive	Correct
90	>100	3.9	No	0.039	Inactive	Correct
86	>100	3.1	No	0.015	Inactive	Correct
126	>100	2.6	No	0.049	Inactive	Correct
133	>100	1.2	No	0.101	Inactive	Correct
125	>100	0.7	No	0.120	Inactive	Correct
139	>100	0.4	No	0.088	Inactive	Correct
146	>100	<0	No	0.104	Inactive	Correct
147	>100	<0	No	0.001	Inactive	Correct
72	>100	<0	No	0.054	Inactive	Correct
69	>100	<0	No	0.307	Inactive	Correct
79	>100	<0	No	0.478	Inactive	Correct

5.10.8 Binary QSAR PC analysis

Descriptor code	1 PC	2 PCs	3 PCs	4 PCs	5 PCs
RMSE	5.0536	5.0345	5.0721	5.0592	5.0480
1-corr	0.1875	0.1023	0.3293	0.2054	0.1311

5.11 MD simulation and analysis

5.11.1 YASARA script - converting YASARA snapshots to .PDB files

Printed below is the file content in which YASARA MD simulation snapshots are converted into PDB files for further analysis with bio3D or visual inspection:

```
OnError Exit
Name = 'AtPARP1_NAD_CNA'
MacroTarget = './(Name)' # define MD simulation to be worked on

for i= 01250 to 01750 # define range of MD to be converted
  LoadSce (MacroTarget) # load MD "template" file,
  LoadSce (MacroTarget)_water # in which the system is defined
  ForceField AMBER03,SetPar=Yes # define FF and parameters
  Interactions Bond,Angle,Dihedral,Planarity,Coulomb,VdW

  # for each snapshot, load it:
  LoadSIM (Name)(i).sim, assignSec = Yes
  # Delres hoh with distance >5 from res Nad
  # DelObj 3
  # DelObj 2
  swapobj 3,2
  DelObj 3 # define protein as chain A
  NameMol res hoh, C # define water as chain B
  NameMol res Nad, B #
  NameMol res Cna, B # define ligands as chain C
  Joinobj 2,1 # combine to single complex, save as PDB
  SavePDB 1, (Name)(i).pdb,Format=PDB,Transform=No
  Clear # clear system and exit when finished
exit
```

5.11.2 R script for preparing MD analysis

```
cd D:\mds\1of3\pdbs\ # change into directory of pdb files
copy /b *.pdb 1of3_all.pdb # concatenate snapshot files into one file

# use catdcd to convert pdb file into dcd trajectory file
catdcd.exe -o D:\mds\1of3\pdbs\1of3_all.dcd -pdb D:\mds\1of3\pdbs\1of3_all.pdb

# read trajectory file into R
dcdfile_1of3 <-"D:\mds\1of3\pdbs\1of3_all.dcd"
dcd_1of3<-read.dcd(dcdfile_1of3) [1:4000,]

# read reference pdb file into R
pdbfile_1of3<-"D:\mds\1of3\pdbs\AtPARP1_NAD_CNA00000.pdb"
pdb_1of3<-read.pdb(pdbfile_1of3)

# get all C-alpha indices
ca_inds_1of3<-atom.select(pdb_1of3,elety="CA")

# superpose all snapshots
fit_1of3<-fit.xyz(fixed=pdb_1of3$xyz, mobile=dcd_1of3, fixed.inds=ca_inds_1of3$xyz,
mobile.inds=ca_inds_1of3$xyz)

#calculate rmsd
rmsd_1of3<-rmsd(fit_1of3[1,ca_inds_1of3$xyz], fit_1of3[,ca_inds_1of3$xyz])
```

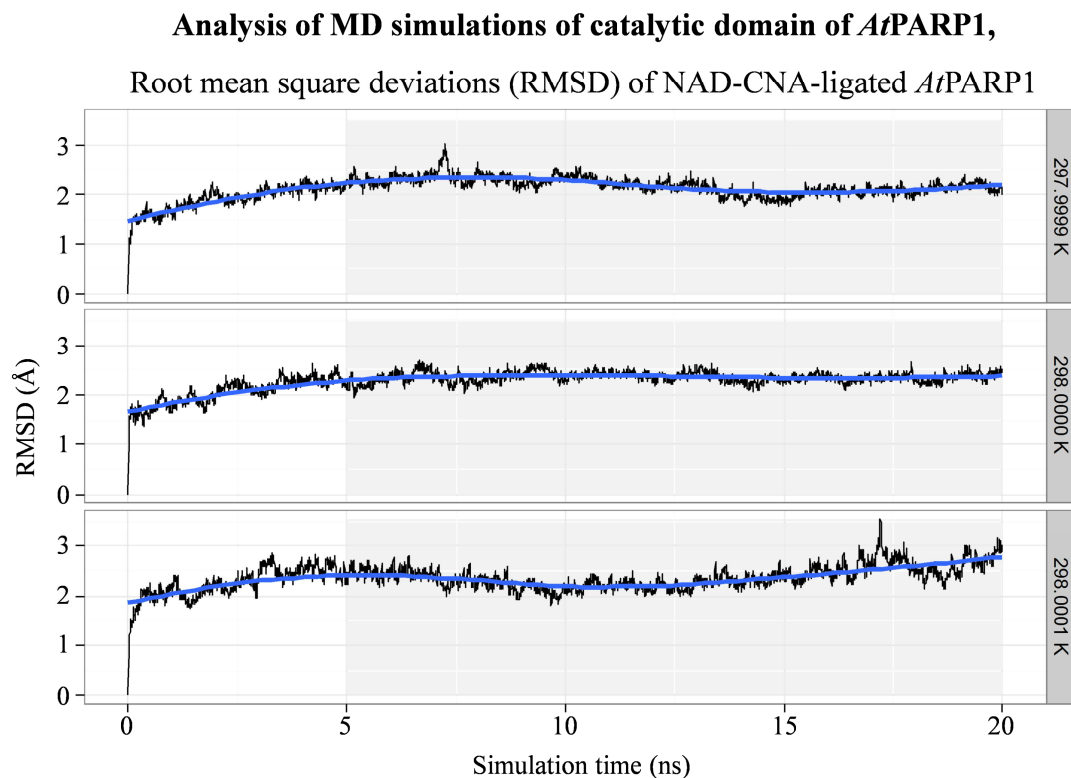
5.11.3 RMSD analysis of NAD-CNA-ligated *At*PARP1

Figure 5.11: Results of MD simulations of NAD-CNA-ligated *At*PARP1. Distributions of differences between calculated CDF and estimated CDF (expressed as e_{CDF}) for all 10 independent docking runs with HsPARP1 inhibitors (A) and HsPARP1 decoys (B).

Analysis of MD simulations of the catalytic domain of *At*PARP1

Binding of adenine moiety of CNA

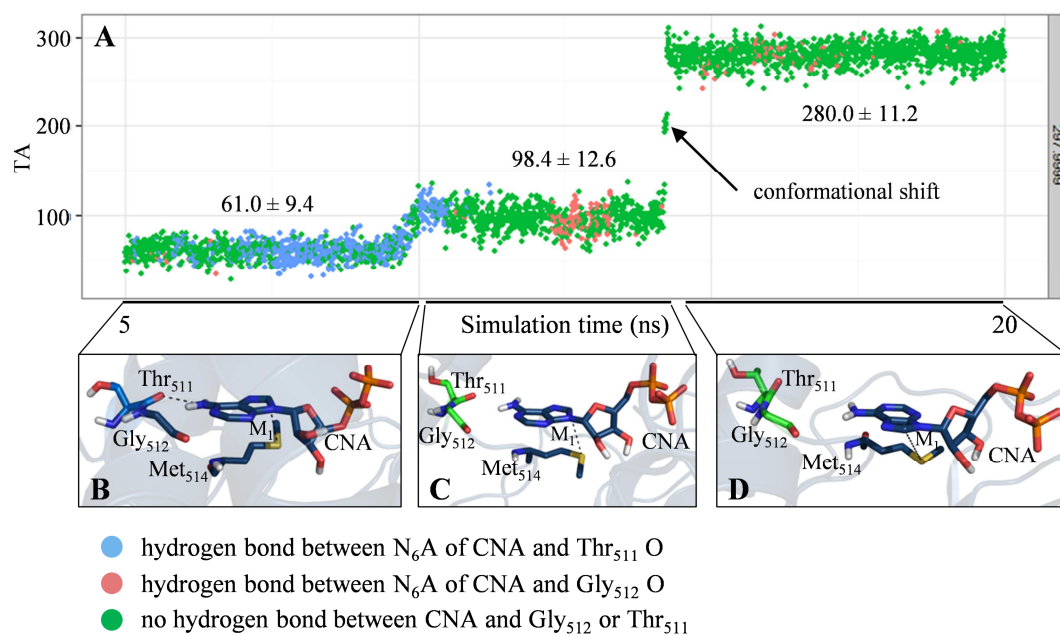


Figure 5.12: Results of MD simulations of NAD-CNA-ligated *At*PARP1. Distributions of differences between calculated CDF and estimated CDF (expressed as e_{CDF}) for all 10 independent docking runs with HsPARP1 inhibitors (A) and HsPARP1 decoys (B).

Analysis of MD simulations of the catalytic domain of *At*PARP1

Binding of adenine moiety of CNA

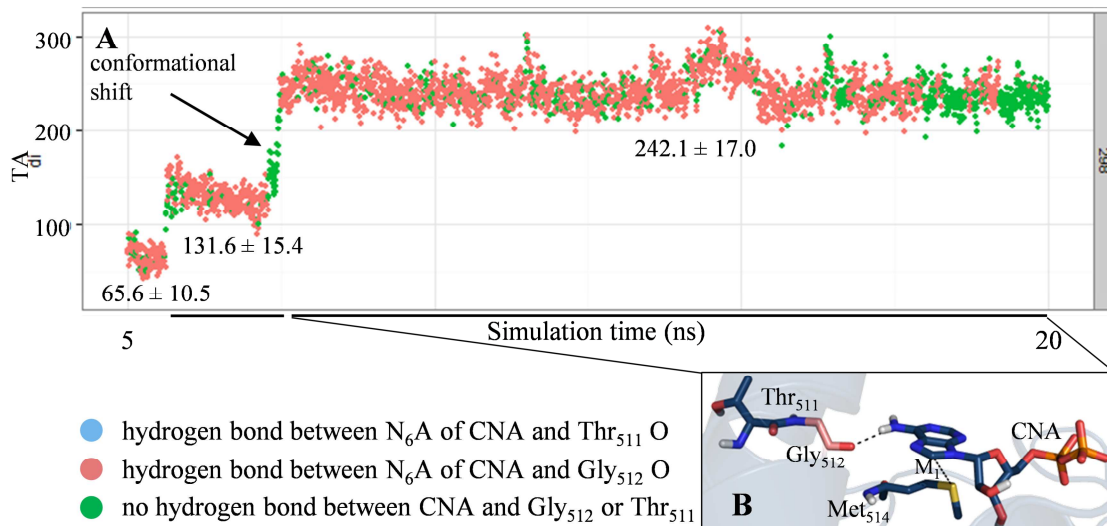


Figure 5.13: Results of MD simulations of NAD-CNA-ligated *At*PARP1. Distributions of differences between calculated CDF and estimated CDF (expressed as e_{CDF}) for all 10 independent docking runs with HsPARP1 inhibitors (A) and HsPARP1 decoys (B).

Analysis of MD simulations of the catalytic domain of *At*PARP1

Binding of adenine moiety of CNA

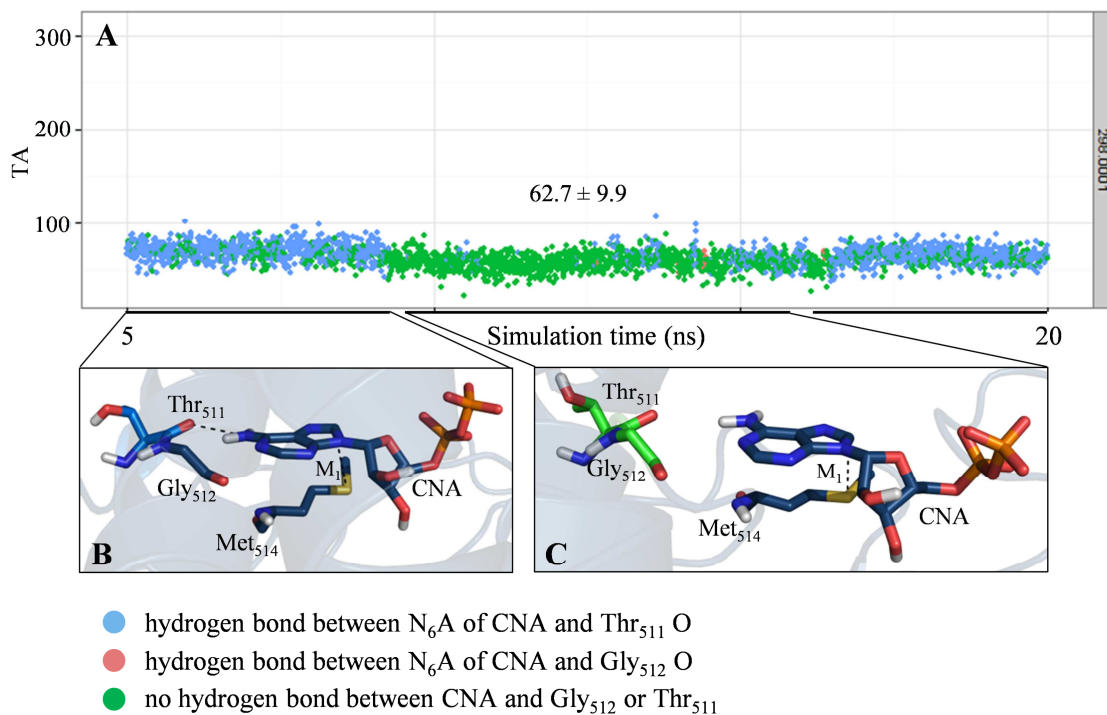


Figure 5.14: Results of MD simulations of NAD-CNA-ligated *At*PARP1. Distributions of differences between calculated CDF and estimated CDF (expressed as e_{CDF}) for all 10 independent docking runs with HsPARP1 inhibitors (A) and HsPARP1 decoys (B).

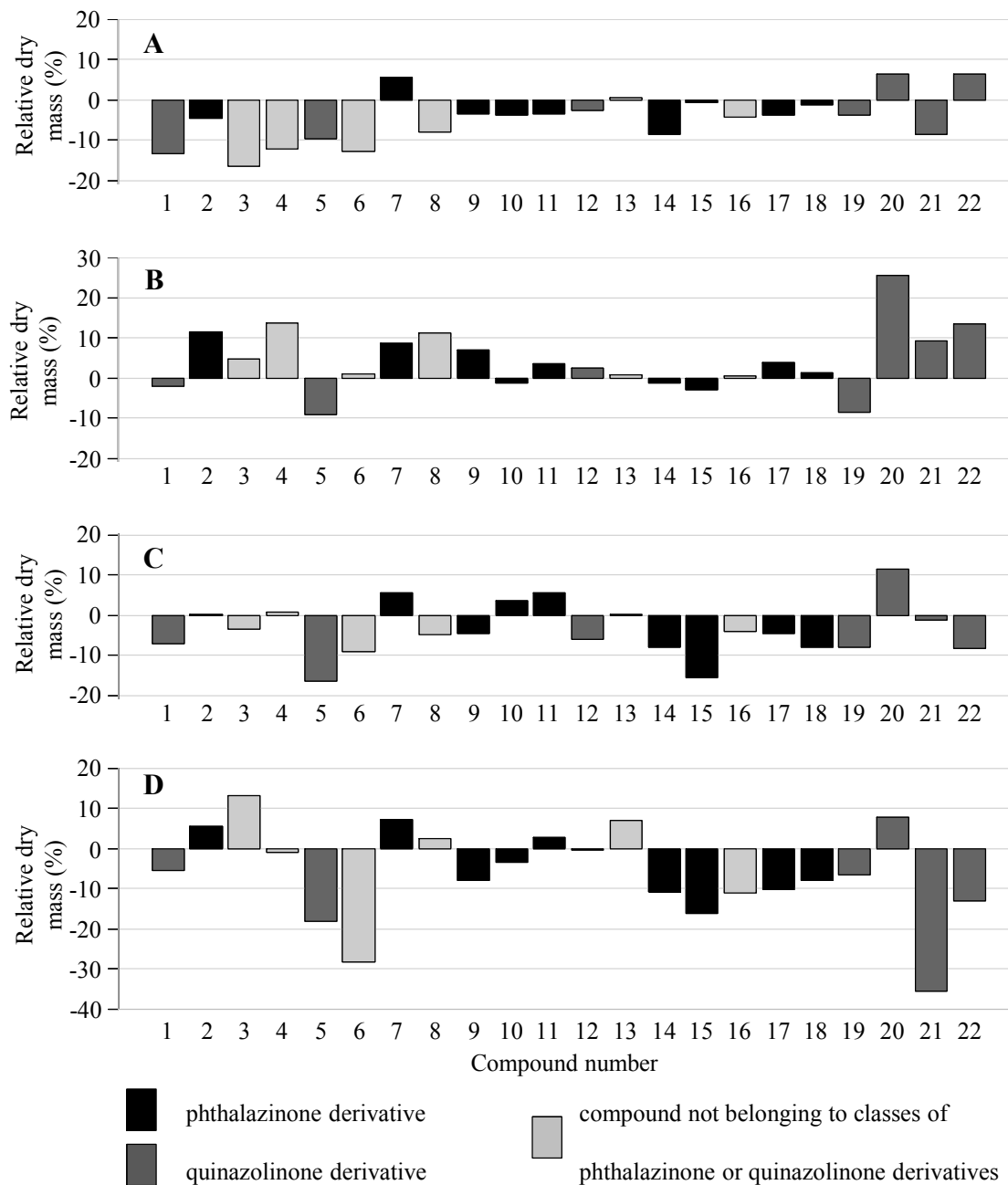
5.12 *Lolium perenne* screening resultsDose-dependent relative dry mass production of *Lolium perenne* shoot, in comparison to stressed and untreated control

Figure 5.15: *Lolium perenne* relative dry mass production for 22 AtPARP1 inhibitors
Four concentrations of 1 μM (A), 10 μM (B), 25 μM (C) and 50 μM (D)

References

- (1) Roitt, I. M. *Biochem. J.* 1956, *63* (2), 300–307.
- (2) Chambon, P.; Weill, J. D.; Mandel, P. *Biochem. Biophys. Res. Commun.* 1963, *11*, 39–43.
- (3) Citarelli, M.; Teotia, S.; Lamb, R. S. *BMC Evol. Biol.* 2010, *10* (1), 308.
- (4) Hassa, P. O.; Haenni, S. S.; Elser, M.; Hottiger, M. O. *Microbiol. Mol. Biol. Rev.* 2006, *70* (3), 789–829.
- (5) Otto, H.; Reche, P. A.; Bazan, F.; Dittmar, K.; Haag, F.; Koch-Nolte, F. *BMC Genomics* 2005, *6* (1), 139.
- (6) Amé, J.-C.; Spenlehauer, C.; de Murcia, G. *BioEssays* 2004, *26* (8), 882–893.
- (7) Hakmé, A.; Wong, H.-K.; Dantzer, F.; Schreiber, V. *EMBO Rep.* 2008, *9* (11), 1094–1100.
- (8) Hassa, P., O. *Front. Biosci.* 2008, *13* (13), 3046.
- (9) Schreiber, V.; Dantzer, F.; Ame, J.-C.; de Murcia, G. *Nat. Rev. Mol. Cell Biol.* 2006, *7* (7), 517–528.
- (10) Kleine, H.; Poreba, E.; Lesniewicz, K.; Hassa, P. O.; Hottiger, M. O.; Litchfield, D. W.; Shilton, B. H.; Lüscher, B. *Mol. Cell* 2008, *32* (1), 57–69.
- (11) D'Amours, D.; Desnoyers, S.; D'Silva, I.; Poirier, G. G. *Mol. Cell* 1999, *39* (1), 249–268.
- (12) Amé, J.-C.; Rolli, V.; Schreiber, V.; Niedergang, C.; Apiou, F.; Decker, P.; Muller, S.; Höger, T.; Murcia, J. M.; Murcia, G. de. *J. Biol. Chem.* 1999, *274* (25), 17860–17868.
- (13) Langelier, M.-F.; Servent, K. M.; Rogers, E. E.; Pascal, J. M. *J. Biol. Chem.* 2008, *283* (7), 4105–4114.
- (14) Langelier, M.-F.; Ruhl, D. D.; Planck, J. L.; Kraus, W. L.; Pascal, J. M. *J. Biol. Chem.* 2010, *285* (24), 18877–18887.
- (15) Desmarais, Y.; Ménard, L.; Lagueux, J.; Poirier, G. G. *Biochim. Biophys. Acta BBA - Protein Struct. Mol. Enzymol.* 1991, *1078* (2), 179–186.
- (16) Rolli, V.; O'Farrell, M.; Ménessier-de Murcia, J.; de Murcia, G. *Biochemistry (Mosc.)* 1997, *36* (40), 12147–12154.
- (17) Marsischky, G. T.; Wilson, B. A.; Collier, R. J. *J. Biol. Chem.* 1995, *270* (7), 3247–3254.
- (18) Krishnakumar, R.; Kraus, W. L. *Mol. Cell* 2010, *39* (1), 8–24.
- (19) Willmitzer, L. *FEBS Lett.* 1979, *108* (1), 13–16.
- (20) Whitby, A. J.; Stone, P. R.; Whish, W. J. D. *Biochem. Biophys. Res. Commun.* 1979, *90* (4), 1295–1304.
- (21) Ahlfors, R.; Lång, S.; Overmyer, K.; Jaspers, P.; Brosché, M.; Tauriainen, A.; Kollist, H.; Tuominen, H.; Belles-Boix, E.; Piippo, M.; Inzé, D.; Palva, E. T.; Kangasjärvi, J. *Plant Cell Online* 2004, *16* (7), 1925–1937.
- (22) Babiychuk, E.; Cottrill, P. B.; Storozhenko, S.; Fuangthong, M.; Chen, Y.; O'Farrell, M. K.; Van Montagu, M.; Inzé, D.; Kushnir, S. *Plant J.* 1998, *15* (5), 635–645.

- (23) Lepiniec, L.; Babiychuk, E.; Kushnir, S.; Van Montagu, M.; Inze', D. *FEBS Lett.* 1995, *364* (2), 103–108.
- (24) Okubo, S.; Hara, F.; Tsuchida, Y.; Shimotakahara, S.; Suzuki, S.; Hatanaka, H.; Yokoyama, S.; Tanaka, H.; Yasuda, H.; Shindo, H. *J. Biol. Chem.* 2004, *279* (30), 31455–31461.
- (25) Ryu, H.; Azuma, Y. *J. Biol. Chem.* 2010, *285* (42), 32576–32585.
- (26) Mahajan, P. B.; Zuo, Z. *Plant Physiol.* 1998, *118* (3), 895–905.
- (27) Becerra, C.; Puigdomenech, P.; Vicient, C. M. *BMC Genomics* 2006, *7* (1), 38.
- (28) Jaspers, P.; Overmyer, K.; Wrzaczek, M.; Vainonen, J. P.; Blomster, T.; Salojärvi, J.; Reddy, R. A.; Kangasjärvi, J. *BMC Genomics* 2010, *11* (1), 170.
- (29) Belles-Boix, E.; Babiychuk, E.; Van Montagu, M.; Inzé, D.; Kushnir, S. *FEBS Lett.* 2000, *482* (1–2), 19–24.
- (30) Jaspers, P.; Blomster, T.; Brosché, M.; Salojärvi, J.; Ahlfors, R.; Vainonen, J. P.; Reddy, R. A.; Immink, R.; Angenent, G.; Turck, F.; Overmyer, K.; Kangasjärvi, J. *Plant J.* 2009, *60* (2), 268–279.
- (31) Ruf, A.; Rolli, V.; de Murcia, G.; Schulz, G. E. *J. Mol. Biol.* 1998, *278* (1), 57–65.
- (32) Scheuring, J.; Schramm, V. L. *Biochemistry (Mosc.)* 1997, *36* (15), 4526–4534.
- (33) Bellocchi, D.; Costantino, G.; Pellicciari, R.; Re, N.; Marrone, A.; Coletti, C. *ChemMedChem* 2006, *1* (5), 533–539.
- (34) Tong, W.-M.; Cortes, U.; Wang, Z.-Q. *Biochim. Biophys. Acta BBA - Rev. Cancer* 2001, *1552* (1), 27–37.
- (35) Rouleau, M.; Aubin, R. A.; Poirier, G. G. *J. Cell Sci.* 2004, *117* (6), 815–825.
- (36) Tulin, A.; Spradling, A. *Science* 2003, *299* (5606), 560–562.
- (37) Leppard, J. B.; Dong, Z.; Mackey, Z. B.; Tomkinson, A. E. *Mol. Cell. Biol.* 2003, *23* (16), 5919–5927.
- (38) Dantzer, F.; de la Rubia, G.; Ménissier-de Murcia, J.; Hostomsky, Z.; de Murcia, G.; Schreiber, V. *Biochemistry (Mosc.)* 2000, *39* (25), 7559–7569.
- (39) Masson, M.; Niedergang, C.; Schreiber, V.; Muller, S.; Murcia, J. M.; Murcia, G. de. *Mol. Cell. Biol.* 1998, *18* (6), 3563–3571.
- (40) Schreiber, V.; Amé, J.-C.; Dollé, P.; Schultz, I.; Rinaldi, B.; Fraulob, V.; Murcia, J. M.; Murcia, G. de. *J. Biol. Chem.* 2002, *277* (25), 23028–23036.
- (41) Hassa, P. O.; Buerki, C.; Lombardi, C.; Imhof, R.; Hottiger, M. O. *J. Biol. Chem.* 2003, *278* (46), 45145–45153.
- (42) Oei, S. L.; Griesenbeck, J.; Schweiger, M.; Ziegler, M. *J. Biol. Chem.* 1998, *273* (48), 31644–31647.
- (43) Simbulan-Rosenthal, C. M.; Rosenthal, D. S.; Hiltz, H.; Hickey, R.; Malkas, L.; Applegren, N.; Wu, Y.; Bers, G.; Smulson, M. E. *Biochemistry (Mosc.)* 1996, *35* (36), 11622–11633.
- (44) Hoeijmakers, J. H. J. *Nature* 2001, *411* (6835), 366–374.
- (45) Caldecott, K. W. *BioEssays* 2001, *23* (5), 447–455.
- (46) Fan, J.; Wilson III, D. M. *Free Radic. Biol. Med.* 2005, *38* (9), 1121–1138.
- (47) Park, C.-J.; Choi, B.-S. *FEBS J.* 2006, *273* (8), 1600–1608.
- (48) Cahill, D. *Front. Biosci.* 2006, *11* (1), 1958.
- (49) Christmann, M.; Tomicic, M. T.; Roos, W. P.; Kaina, B. *Toxicology* 2003, *193* (1–2), 3–34.

- (50) Caldecott, K. W. *DNA Repair* 2003, 2 (9), 955–969.
- (51) Vodenicharov, M. D.; Sallmann, F. R.; Satoh, M. S.; Poirier, G. G. *Nucleic Acids Res.* 2000, 28 (20), 3887–3896.
- (52) Shall, S.; de Murcia, G. *Mutat. Res. Repair* 2000, 460 (1), 1–15.
- (53) Caldecott, K. W.; Aoufouchi, S.; Johnson, P.; Shall, S. *Nucleic Acids Res.* 1996, 24 (22), 4387–4394.
- (54) Prasad, R.; Lavrik, O. I.; Kim, S.-J.; Kedar, P.; Yang, X.-P.; Berg, B. J. V.; Wilson, S. H. *J. Biol. Chem.* 2001, 276 (35), 32411–32414.
- (55) Parsons, J. L.; Dianov, G. L. *Biochem. Soc. Trans.* 2004, 32 (Pt 6), 962–963.
- (56) Hartwell, L. H.; Szankasi, P.; Roberts, C. J.; Murray, A. W.; Friend, S. H. *Science* 1997, 278 (5340), 1064–1068.
- (57) Bryant, H. E.; Schultz, N.; Thomas, H. D.; Parker, K. M.; Flower, D.; Lopez, E.; Kyle, S.; Meuth, M.; Curtin, N. J.; Helleday, T. *Nature* 2005, 434 (7035), 913–917.
- (58) Farmer, H.; McCabe, N.; Lord, C. J.; Tutt, A. N. J.; Johnson, D. A.; Richardson, T. B.; Santarosa, M.; Dillon, K. J.; Hickson, I.; Knights, C.; Martin, N. M. B.; Jackson, S. P.; Smith, G. C. M.; Ashworth, A. *Nature* 2005, 434 (7035), 917–921.
- (59) Wooster, R.; Weber, B. L. *N. Engl. J. Med.* 2003, 348 (23), 2339–2347.
- (60) Gudmundsdottir, K.; Ashworth, A. *Oncogene* 2006, 25 (43), 5864–5874.
- (61) Lord, C. J.; Ashworth, A. *Curr. Opin. Pharmacol.* 2008, 8 (4), 363–369.
- (62) Saffhill, R.; Ockey, C. H. *Chromosoma* 1985, 92 (3), 218–224.
- (63) Pieper, A. A.; Verma, A.; Zhang, J.; Snyder, S. H. *Trends Pharmacol. Sci.* 1999, 20 (4), 171–181.
- (64) Purnell, M. R.; Whish, W. J. *Biochem. J.* 1980, 185 (3), 775–777.
- (65) Durkacz, B. W.; Omidiji, O.; Gray, D. A.; Shall, S. *Nature* 1980, 283 (5747), 593–596.
- (66) Ferraris, D. V. *J. Med. Chem.* 2010, 53 (12), 4561–4584.
- (67) Watson, C. Y.; Whish, W. J. D.; Threadgill, M. D. *Bioorg. Med. Chem.* 1998, 6 (6), 721–734.
- (68) Banasik, M.; Komura, H.; Shimoyama, M.; Ueda, K. *J. Biol. Chem.* 1992, 267 (3), 1569–1575.
- (69) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. *Nucleic Acids Res.* 2000, 28 (1), 235–242.
- (70) Kinoshita, T.; Nakanishi, I.; Warizaya, M.; Iwashita, A.; Kido, Y.; Hattori, K.; Fujii, T. *FEBS Lett.* 2004, 556 (1–3), 43–46.
- (71) Karlberg, T.; Hammarström, M.; Schütz, P.; Svensson, L.; Schüler, H. *Biochemistry (Mosc.)* 2010, 49 (6), 1056–1058.
- (72) Ruf, A.; de Murcia, G.; Schulz, G. E. *Biochemistry (Mosc.)* 1998, 37 (11), 3893–3900.
- (73) Miknyoczki, S. J.; Jones-Bolin, S.; Pritchard, S.; Hunter, K.; Zhao, H.; Wan, W.; Ator, M.; Bihovsky, R.; Hudkins, R.; Chatterjee, S.; Klein-Szanto, A.; Dionne, C.; Ruggeri, B. *Mol. Cancer Ther.* 2003, 2 (4), 371–382.
- (74) Tentori, L.; Leonetti, C.; Scarsella, M.; d' Amati, G.; Vergati, M.; Portarena, I.; Xu, W.; Kalish, V.; Zupi, G.; Zhang, J.; Graziani, G. *Clin. Cancer Res.* 2003, 9 (14), 5370–5379.

- (75) Calabrese, C. R.; Almassy, R.; Barton, S.; Batey, M. A.; Calvert, A. H.; Canan-Koch, S.; Durkacz, B. W.; Hostomsky, Z.; Kumpf, R. A.; Kyle, S.; Li, J.; Maegley, K.; Newell, D. R.; Notarianni, E.; Stratford, I. J.; Skalitzky, D.; Thomas, H. D.; Wang, L.-Z.; Webber, S. E.; Williams, K. J.; Curtin, N. J. *J. Natl. Cancer Inst.* 2004, *96* (1), 56–67.
- (76) Tentori, L.; Graziani, G. *Pharmacol. Res.* 2005, *52* (1), 25–33.
- (77) List of Clinical Trials in connection with the keyword PARP <http://clinicaltrials.gov/ct2/results?term=PARP&Search=Search> (accessed Dec 29, 2014).
- (78) Plummer, R.; Jones, C.; Middleton, M.; Wilson, R.; Evans, J.; Olsen, A.; Curtin, N.; Boddy, A.; McHugh, P.; Newell, D.; Harris, A.; Johnson, P.; Steinfeldt, H.; Dewji, R.; Wang, D.; Robson, L.; Calvert, H. *Clin. Cancer Res.* 2008, *14* (23), 7917–7923.
- (79) Thomas, H. D.; Calabrese, C. R.; Batey, M. A.; Canan, S.; Hostomsky, Z.; Kyle, S.; Maegley, K. A.; Newell, D. R.; Skalitzky, D.; Wang, L.-Z.; Webber, S. E.; Curtin, N. J. *Mol. Cancer Ther.* 2007, *6* (3), 945–956.
- (80) Plummer, R.; Lorigan, P.; Steven, N.; Scott, L.; Middleton, M. R.; Wilson, R. H.; Mulligan, E.; Curtin, N.; Wang, D.; Dewji, R.; Abbattista, A.; Gallo, J.; Calvert, H. *Cancer Chemother. Pharmacol.* 2013, *71* (5), 1191–1199.
- (81) Miknyoczki, S.; Chang, H.; Grobelny, J.; Pritchard, S.; Worrell, C.; McGann, N.; Ator, M.; Husten, J.; Deibold, J.; Hudkins, R.; Zulli, A.; Parchment, R.; Ruggeri, B. *Mol. Cancer Ther.* 2007, *6* (8), 2290–2302.
- (82) Loh Jr., V. M.; Cockcroft, X.; Dillon, K. J.; Dixon, L.; Drzewiecki, J.; Eversley, P. J.; Gomez, S.; Hoare, J.; Kerrigan, F.; Matthews, I. T. W.; Menear, K. A.; Martin, N. M. B.; Newton, R. F.; Paul, J.; Smith, G. C. M.; Vile, J.; Whittle, A. J. *Bioorg. Med. Chem. Lett.* 2005, *15* (9), 2235–2238.
- (83) Cockcroft, X.; Dillon, K. J.; Dixon, L.; Drzewiecki, J.; Kerrigan, F.; Loh Jr., V. M.; Martin, N. M. B.; Menear, K. A.; Smith, G. C. M. *Bioorg. Med. Chem. Lett.* 2006, *16* (4), 1040–1044.
- (84) Menear, K. A.; Adcock, C.; Alonso, F. C.; Blackburn, K.; Copsey, L.; Drzewiecki, J.; Fundo, A.; Le Gall, A.; Gomez, S.; Javaid, H.; Lence, C. F.; Martin, N. M. B.; Mydlowski, C.; Smith, G. C. M. *Bioorg. Med. Chem. Lett.* 2008, *18* (14), 3942–3945.
- (85) Sakai, W.; Swisher, E. M.; Karlan, B. Y.; Agarwal, M. K.; Higgins, J.; Friedman, C.; Villegas, E.; Jacquemont, C.; Farrugia, D. J.; Couch, F. J.; Urban, N.; Taniguchi, T. *Nature* 2008, *451* (7182), 1116–1120.
- (86) Barber, L. J.; Sandhu, S.; Chen, L.; Campbell, J.; Kozarewa, I.; Fenwick, K.; Assiotis, I.; Rodrigues, D. N.; Reis-Filho, J. S.; Moreno, V.; Mateo, J.; Molife, L. R.; De Bono, J.; Kaye, S.; Lord, C. J.; Ashworth, A. *J. Pathol.* 2013, *229* (3), 422–429.
- (87) Liu, X.; Shi, Y.; Maag, D. X.; Palma, J. P.; Patterson, M. J.; Ellis, P. A.; Surber, B. W.; Ready, D. B.; Soni, N. B.; Lador, U. S.; Xu, A. J.; Iyer, R.; Harlan, J. E.; Solomon, L. R.; Donawho, C. K.; Penning, T. D.; Johnson, E. F.; Shoemaker, A. R. *Clin. Cancer Res.* 2012, *18* (2), 510–523.
- (88) Patel, A. G.; Lorenzo, S. B. D.; Flatten, K. S.; Poirier, G. G.; Kaufmann, S. H. *Clin. Cancer Res.* 2012, *18* (6), 1655–1662.

- (89) Torsoli, A.; Serafino, P. Sanofi Ends Iniparib Research, Plans \$285 Million Charge <http://www.bloomberg.com/news/2013-06-03/sanofi-ends-iniparib-research-plans-285-million-charge.html> (accessed Nov 18, 2013).
- (90) Donawho, C. K.; Luo, Y.; Luo, Y.; Penning, T. D.; Bauch, J. L.; Bouska, J. J.; Bontcheva-Diaz, V. D.; Cox, B. F.; DeWeese, T. L.; Dillehay, L. E.; Ferguson, D. C.; Ghoreishi-Haack, N. S.; Grimm, D. R.; Guan, R.; Han, E. K.; Holley-Shanks, R. R.; Hristov, B.; Idler, K. B.; Jarvis, K.; Johnson, E. F.; Kleinberg, L. R.; Klinghofer, V.; Lasko, L. M.; Liu, X.; Marsh, K. C.; McGonigal, T. P.; Meulbroek, J. A.; Olson, A. M.; Palma, J. P.; Rodriguez, L. E.; Shi, Y.; Stavropoulos, J. A.; Tsurutani, A. C.; Zhu, G.-D.; Rosenberg, S. H.; Giranda, V. L.; Frost, D. J. *Clin. Cancer Res.* 2007, *13* (9), 2728–2737.
- (91) Kummar, S.; Ji, J.; Morgan, R.; Lenz, H.-J.; Puhalla, S. L.; Belani, C. P.; Gandara, D. R.; Allen, D.; Kiesel, B.; Beumer, J. H.; Newman, E. M.; Rubinstein, L.; Chen, A.; Zhang, Y.; Wang, L.; Kinders, R. J.; Parchment, R. E.; Tomaszewski, J. E.; Doroshow, J. H. *Clin. Cancer Res.* 2012, *18* (6), 1726–1734.
- (92) Doucet-Chabeaud, G.; Godon, C.; Brutesco, C.; Murcia, G. de; Kazmaier, M. *Mol. Genet. Genomics* 2001, *265* (6), 954–963.
- (93) Amor, Y.; Babiychuk, E.; Inzé, D.; Levine, A. *FEBS Lett.* 1998, *440* (1–2), 1–7.
- (94) Ramirez-Parra, E.; Gutierrez, C. *Plant Physiol.* 2007, *144* (1), 105–120.
- (95) Zhu, Y.; Weng, M.; Yang, Y.; Zhang, C.; Li, Z.; Shen, W.-H.; Dong, A. *Plant J.* 2011, *66* (3), 443–455.
- (96) Culligan, K. M.; Robertson, C. E.; Foreman, J.; Doerner, P.; Britt, A. B. *Plant J.* 2006, *48* (6), 947–961.
- (97) Chen, I.-P.; Haehnel, U.; Altschmied, L.; Schubert, I.; Puchta, H. *Plant J.* 2003, *35* (6), 771–786.
- (98) Ogawa, T.; Ishikawa, K.; Harada, K.; Fukusaki, E.; Yoshimura, K.; Shigeoka, S. *Plant J.* 2009, *57* (2), 289–301.
- (99) De Block, M.; Verduyn, C.; Brouwer, D. D.; Cornelissen, M. *Plant J.* 2005, *41* (1), 95–106.
- (100) Vanderauwera, S.; Block, M. D.; Steene, N. V. de; Cotte, B. van de; Metzlaff, M.; Breusegem, F. V. *Proc. Natl. Acad. Sci.* 2007, *104* (38), 15150–15155.
- (101) Wu, Y.; Kuzma, J.; Maréchal, E.; Graeff, R.; Lee, H. C.; Foster, R.; Chua, N.-H. *Science* 1997, *278* (5346), 2126–2130.
- (102) Sánchez, J.-P.; Duque, P.; Chua, N.-H. *Plant J.* 2004, *38* (3), 381–395.
- (103) Andrabi, S. A.; Kim, N. S.; Yu, S.-W.; Wang, H.; Koh, D. W.; Sasaki, M.; Klaus, J. A.; Otsuka, T.; Zhang, Z.; Koehler, R. C.; Hurn, P. D.; Poirier, G. G.; Dawson, V. L.; Dawson, T. M. *Proc. Natl. Acad. Sci.* 2006, *103* (48), 18308–18313.
- (104) Hanai, S.; Kanai, M.; Ohashi, S.; Okamoto, K.; Yamada, M.; Takahashi, H.; Miwa, M. *Proc. Natl. Acad. Sci.* 2004, *101* (1), 82–86.
- (105) Fisher, A. E. O.; Hochegger, H.; Takeda, S.; Caldecott, K. W. *Mol. Cell. Biol.* 2007, *27* (15), 5597–5605.
- (106) Koh, D. W.; Lawler, A. M.; Poitras, M. F.; Sasaki, M.; Wattler, S.; Nehls, M. C.; Stöger, T.; Poirier, G. G.; Dawson, V. L.; Dawson, T. M. *Proc. Natl. Acad. Sci. U. S. A.* 2004, *101* (51), 17699–17704.
- (107) Panda, S.; Poirier, G. G.; Kay, S. A. *Dev. Cell* 2002, *3* (1), 51–61.

- (108) Ishikawa, K.; Yoshimura, K.; Ogawa, T.; Shigeoka, S. *Plant Signal. Behav.* 2010, 5 (7), 839–841.
- (109) Ogawa, T.; Ueda, Y.; Yoshimura, K.; Shigeoka, S. *J. Biol. Chem.* 2005, 280 (26), 25277–25283.
- (110) Briggs, A. G.; Bent, A. F. *Trends Plant Sci.* 2011, 16 (7), 372–380.
- (111) Ishikawa, K.; Ogawa, T.; Hirose, E.; Nakayama, Y.; Harada, K.; Fukusaki, E.; Yoshimura, K.; Shigeoka, S. *Plant Physiol.* 2009, 151 (2), 741–754.
- (112) Ge, X.; Li, G.-J.; Wang, S.-B.; Zhu, H.; Zhu, T.; Wang, X.; Xia, Y. *Plant Physiol.* 2007, 145 (1), 204–215.
- (113) Adams-Phillips, L.; Briggs, A. G.; Bent, A. F. *Plant Physiol.* 2010, 152 (1), 267–280.
- (114) Bartsch, M.; Gobbato, E.; Bednarek, P.; Debey, S.; Schultze, J. L.; Bautor, J.; Parker, J. E. *Plant Cell Online* 2006, 18 (4), 1038–1051.
- (115) Jambunathan, N.; Mahalingam, R. *Planta* 2006, 224 (1), 1–11.
- (116) McKiernan, D. P. J. *Drugs* 2006, 66 (6), 743–750.
- (117) Phornphutkul, C.; Introne, W. J.; Perry, M. B.; Bernardini, I.; Murphey, M. D.; Fitzpatrick, D. L.; Anderson, P. D.; Huizing, M.; Anikster, Y.; Gerber, L. H.; Gahl, W. A. *N. Engl. J. Med.* 2002, 347 (26), 2111–2121.
- (118) Dillon, K. J.; Smith, G. C. M.; Martin, N. M. B. *J. Biomol. Screen.* 2003, 8 (3), 347–352.
- (119) Domenighini, M.; Montecucco, C.; Ripka, W. C.; Rappuoli, R. *Mol. Microbiol.* 1991, 5 (1), 23–31.
- (120) Domenighini, M.; Magagnoli, C.; Pizza, M.; Rappuoli, R. *Mol. Microbiol.* 1994, 14 (1), 41–50.
- (121) Bell, C. E.; Eisenberg, D. *Biochemistry (Mosc.)* 1996, 35 (4), 1137–1149.
- (122) Ruf, A.; Murcia, J. M. de; Murcia, G. de; Schulz, G. E. *Proc. Natl. Acad. Sci.* 1996, 93 (15), 7481–7485.
- (123) Jørgensen, R.; Wang, Y.; Visschedyk, D.; Merrill, A. R. *EMBO Rep.* 2008, 9 (8), 802–809.
- (124) O’Neal, C. J.; Jobling, M. G.; Holmes, R. K.; Hol, W. G. J. *Science* 2005, 309 (5737), 1093–1096.
- (125) Lee, Y.-M.; Babu, C. S.; Chen, Y. C.; Milcic, M.; Qu, Y.; Lim, C. *J. Med. Chem.* 2010, 53 (10), 4038–4049.
- (126) García-Sosa, A. T.; Firth-Clark, S.; Mancera, R. L. *J. Chem. Inf. Model.* 2005, 45 (3), 624–633.
- (127) Costantino, G.; Macchiarulo, A.; Camaioni, E.; Pellicciari, R. *J. Med. Chem.* 2001, 44 (23), 3786–3794.
- (128) Riahi, S.; Pourbasheer, E.; Dinarvand, R.; Ganjali, M. R.; Norouzi, P. *Chem. Biol. Drug Des.* 2008, 72 (6), 575–584.
- (129) Cramer, R. D.; Patterson, D. E.; Bunce, J. D. *J. Am. Chem. Soc.* 1988, 110 (18), 5959–5967.
- (130) Klebe, G.; Abraham, U.; Mietzner, T. *J. Med. Chem.* 1994, 37 (24), 4130–4146.
- (131) Zeng, H.; Zhang, H.; Jang, F.; Zhao, L.; Zhang, J. *Chem. Biol. Drug Des.* 2011, 78 (3), 333–352.
- (132) Rewatkar, P. V.; Kokil, G. R.; Raut, M. K. *Med. Chem. Res.* 2011, 20 (7), 877–886.

- (133) Pellicciari, R.; Camaioni, E.; Costantino, G.; Marinozzi, M.; Macchiarulo, A.; Moroni, F.; Natalini, B. *Il Farm.* 2003, 58 (9), 851–858.
- (134) Ishida, J.; Yamamoto, H.; Kido, Y.; Kamijo, K.; Murano, K.; Miyake, H.; Ohkubo, M.; Kinoshita, T.; Warizaya, M.; Iwashita, A.; Mihara, K.; Matsuoka, N.; Hattori, K. *Bioorg. Med. Chem.* 2006, 14 (5), 1378–1390.
- (135) Oliver, A. W.; Amé, J.-C.; Roe, S. M.; Good, V.; Murcia, G. de; Pearl, L. H. *Nucleic Acids Res.* 2004, 32 (2), 456–464.
- (136) Novikov, F. N.; Stroylov, V. S.; Stroganov, O. V.; Kulkov, V.; Chilov, G. G. *J. Mol. Model.* 2009, 15 (11), 1337–1347.
- (137) Stroganov, O. V.; Novikov, F. N.; Stroylov, V. S.; Kulkov, V.; Chilov, G. G. *J. Chem. Inf. Model.* 2008, 48 (12), 2371–2385.
- (138) Korb, O.; Monecke, P.; Hessler, G.; Stütze, T.; Exner, T. E. *J. Chem. Inf. Model.* 2010, 50 (9), 1669–1681.
- (139) Ott, K.-H.; Arani, N.; Singh, B.; Stockton, G. W. *Phytochemistry* 2003, 62 (6), 971–985.
- (140) Menne, H. Homepage of Herbicide Resistance Action Committee (HRAC) <http://www.hracglobal.com/Education/ClassificationofHerbicideSiteofAction.aspx>.
- (141) Chipman, D.; Barak, Z.; Schloss, J. V. *Biochim. Biophys. Acta BBA - Protein Struct. Mol. Enzymol.* 1998, 1385 (2), 401–419.
- (142) Chaleff, R. S.; Mauvais, C. J. *Science* 1984, 224 (4656), 1443–1445.
- (143) Pang, S. S.; Duggleby, R. G.; Guddat, L. W. *J. Mol. Biol.* 2002, 317 (2), 249–262.
- (144) Duggleby, R. G.; McCourt, J. A.; Guddat, L. W. *Plant Physiol. Biochem.* 2008, 46 (3), 309–324.
- (145) Wang, J.-G.; Xiao, Y.-J.; Li, Y.-H.; Ma, Y.; Li, Z.-M. *Bioorg. Med. Chem.* 2007, 15 (1), 374–380.
- (146) Shang, J.; Wang, W.-M.; Li, Y.-H.; Song, H.-B.; Li, Z.-M.; Wang, J.-G. *J. Agric. Food Chem.* 2012, 60 (34), 8286–8293.
- (147) Wang, J.; Tan, H.; Li, Y.; Ma, Y.; Li, Z.; Guddat, L. W. *J. Agric. Food Chem.* 2011, 59 (18), 9892–9900.
- (148) Ackerman, F. *Int. J. Occup. Environ. Health* 2007, 13 (4), 437–445.
- (149) Tillitt, D. E.; Papoulias, D. M.; Whyte, J. J.; Richter, C. A. *Aquat. Toxicol.* 2010, 99 (2), 149–159.
- (150) Jalaie, M.; Erickson, J. A. *J. Comput. Aided Mol. Des.* 2000, 14 (2), 181–197.
- (151) Bordás, B.; Kőmíves, T.; Lopata, A. *Pest Manag. Sci.* 2003, 59 (4), 393–400.
- (152) Key Organics <http://www.keyorganics.net/>.
- (153) Key Organics Agromediate <http://www.keyorganics.net/services/bionet-products/agromediate/>.
- (154) Bellocchi, D.; Macchiarulo, A.; Costantino, G.; Pellicciari, R. *Bioorg. Med. Chem.* 2005, 13 (4), 1151–1157.
- (155) Dunn, D.; Husten, J.; Ator, M. A.; Chatterjee, S. *Bioorg. Med. Chem. Lett.* 2007, 17 (2), 542–545.
- (156) Hattori, K.; Kido, Y.; Yamamoto, H.; Ishida, J.; Iwashita, A.; Mihara, K. *Bioorg. Med. Chem. Lett.* 2007, 17 (20), 5577–5581.

- (157) Hattori, K.; Kido, Y.; Yamamoto, H.; Ishida, J.; Kamijo, K.; Murano, K.; Ohkubo, M.; Kinoshita, T.; Iwashita, A.; Mihara, K.; Yamazaki, S.; Matsuoka, N.; Teramura, Y.; Miyake, H. *J. Med. Chem.* 2004, *47* (17), 4151–4154.
- (158) Huang, N.; Shoichet, B. K.; Irwin, J. J. *J. Med. Chem.* 2006, *49* (23), 6789–6801.
- (159) EMBOSS Needle http://www.ebi.ac.uk/Tools/psa/emboss_needle/.
- (160) Needleman, S. B.; Wunsch, C. D. *J. Mol. Biol.* 1970, *48* (3), 443–453.
- (161) Clustal Omega <http://www.ebi.ac.uk/Tools/msa/clustalo/>.
- (162) Goujon, M.; McWilliam, H.; Li, W.; Valentin, F.; Squizzato, S.; Paern, J.; Lopez, R. *Nucleic Acids Res.* 2010, *38* (suppl 2), W695–W699.
- (163) Sievers, F.; Wilm, A.; Dineen, D.; Gibson, T. J.; Karplus, K.; Li, W.; Lopez, R.; McWilliam, H.; Remmert, M.; Söding, J.; Thompson, J. D.; Higgins, D. G. *Mol. Syst. Biol.* 2011, *7* (1).
- (164) Saitou, N.; Nei, M. *Mol. Biol. Evol.* 1987, *4* (4), 406–425.
- (165) *Molecular Operating Environment (MOE)*; Chemical Computing Group Inc.: 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7, 2013, 2008.
- (166) *Molecular Operating Environment (MOE)*; Chemical Computing Group Inc.: 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7, 2013, 2009.
- (167) *Molecular Operating Environment (MOE)*; Chemical Computing Group Inc.: 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7, 2013, 2010.
- (168) *Molecular Operating Environment (MOE)*; Chemical Computing Group Inc.: 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7, 2011, 2011.
- (169) *Molecular Operating Environment (MOE)*; Chemical Computing Group Inc.: 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7, 2013, 2012.
- (170) Labute, P. *J. Chem. Inf. Model.* 2010, *50* (5), 792–800.
- (171) Hawkins, P. C. D.; Skillman, A. G.; Warren, G. L.; Ellingson, B. A.; Stahl, M. T. *J. Chem. Inf. Model.* 2010, *50* (4), 572–584.
- (172) Mills, J. E. J.; Dean, P. M. *J. Comput. Aided Mol. Des.* 1996, *10* (6), 607–622.
- (173) Wlodek, S.; Skillman, A. G.; Nicholls, A. *Acta Crystallogr. D Biol. Crystallogr.* 2006, *62* (7), 741–749.
- (174) Krieger, E.; Joo, K.; Lee, J.; Lee, J.; Raman, S.; Thompson, J.; Tyka, M.; Baker, D.; Karplus, K. *Proteins Struct. Funct. Bioinforma.* 2009, *77* (S9), 114–122.
- (175) Krieger, E.; Darden, T.; Nabuurs, S. B.; Finkelstein, A.; Vriend, G. *Proteins Struct. Funct. Bioinforma.* 2004, *57* (4), 678–683.
- (176) 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.; Kollman, P. *J. Comput. Chem.* 2003, *24* (16), 1999–2012.
- (177) King, R. D.; Sternberg, M. J. E. *Protein Sci.* 1996, *5* (11), 2298–2310.
- (178) Krieger, E.; Nielsen, J. E.; Spronk, C. A. E. M.; Vriend, G. *J. Mol. Graph. Model.* 2006, *25* (4), 481–486.
- (179) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. *J. Chem. Phys.* 1984, *81* (8), 3684–3690.
- (180) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *J. Chem. Phys.* 1995, *103* (19), 8577.
- (181) Watts, K. S.; Dalal, P.; Murphy, R. B.; Sherman, W.; Friesner, R. A.; Shelley, J. C. *J. Chem. Inf. Model.* 2010, *50* (4), 534–546.

- (182) Chen, I.-J.; Foloppe, N. *J. Chem. Inf. Model.* 2010, *50* (5), 822–839.
- (183) Banks, J. L.; Beard, H. S.; Cao, Y.; Cho, A. E.; Damm, W.; Farid, R.; Felts, A. K.; Halgren, T. A.; Mainz, D. T.; Maple, J. R.; Murphy, R.; Philipp, D. M.; Repasky, M. P.; Zhang, L. Y.; Berne, B. J.; Friesner, R. A.; Gallicchio, E.; Levy, R. M. *J. Comput. Chem.* 2005, *26* (16), 1752–1780.
- (184) *Ligprep, version 2.4*; Ligprep, version 2.4, Schrödinger, LLC, New York, NY, 2010.: New York, NY, 2010.
- (185) Shelley, J. C.; Cholleti, A.; Frye, L. L.; Greenwood, J. R.; Timlin, M. R.; Uchimaya, M. *J. Comput. Aided Mol. Des.* 2007, *21* (12), 681–691.
- (186) Greenwood, J. R.; Calkins, D.; Sullivan, A. P.; Shelley, J. C. *J. Comput. Aided Mol. Des.* 2010, *24* (6-7), 591–604.
- (187) Leung, S. S. F.; Mijalkovic, J.; Borrelli, K.; Jacobson, M. P. *J. Chem. Inf. Model.* 2012, *52* (6), 1621–1636.
- (188) Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S. *J. Med. Chem.* 2004, *47* (7), 1739–1749.
- (189) Halgren, T. A.; Murphy, R. B.; Friesner, R. A.; Beard, H. S.; Frye, L. L.; Pollard, W. T.; Banks, J. L. *J. Med. Chem.* 2004, *47* (7), 1750–1759.
- (190) *Glide, version 5.6*; Glide, version 5.6, Schrödinger, LLC, New York, NY, 2010.: New York, NY.
- (191) Eldridge, M. D.; Murray, C. W.; Auton, T. R.; Paolini, G. V.; Mee, R. P. *J. Comput. Aided Mol. Des.* 1997, *11* (5), 425–445.
- (192) Friesner, R. A.; Murphy, R. B.; Repasky, M. P.; Frye, L. L.; Greenwood, J. R.; Halgren, T. A.; Sanschagrin, P. C.; Mainz, D. T. *J. Med. Chem.* 2006, *49* (21), 6177–6196.
- (193) Jones, G.; Willett, P.; Glen, R. C. *J. Mol. Biol.* 1995, *245* (1), 43–53.
- (194) Jones, G.; Willett, P. *Curr. Opin. Biotechnol.* 1995, *6* (6), 652–656.
- (195) Davis, L. *Handbook of Genetic Algorithms*; Van Nostrand Reinhold, 1991.
- (196) Brodmeier, T.; Pretsch, E. *J. Comput. Chem.* 1994, *15* (6), 588–595.
- (197) Clark, K. P.; Ajay. *J. Comput. Chem.* 1995, *16* (10), 1210–1226.
- (198) Mooij, W. T. M.; Verdonk, M. L. *Proteins Struct. Funct. Bioinforma.* 2005, *61* (2), 272–287.
- (199) Murray, C. W.; Auton, T. R.; Eldridge, M. D. *J. Comput. Aided Mol. Des.* 1998, *12* (5), 503–519.
- (200) Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. *Development and Validation of a Genetic Algorithm for Flexible Docking*.
- (201) Korb, O.; Stützel, T.; Exner, T. E. *J. Chem. Inf. Model.* 2009, *49* (1), 84–96.
- (202) Gehlhaar, D. K.; Verkhivker, G. M.; Rejto, P. A.; Sherman, C. J.; Fogel, D. R.; Fogel, L. J.; Freer, S. T. *Chem. Biol.* 1995, *2* (5), 317–324.
- (203) Verkhivker, G. M. *J. Mol. Graph. Model.* 2004, *22* (5), 335–348.
- (204) Verdonk, M. L.; Berdini, V.; Hartshorn, M. J.; Mooij, W. T. M.; Murray, C. W.; Taylor, R. D.; Watson, P. *J. Chem. Inf. Comput. Sci.* 2004, *44* (3), 793–806.
- (205) Verdonk, M. L.; Cole, J. C.; Hartshorn, M. J.; Murray, C. W.; Taylor, R. D. *Proteins Struct. Funct. Bioinforma.* 2003, *52* (4), 609–623.

- (206) Clark, M.; Cramer, R. D.; Van Opdenbosch, N. *J. Comput. Chem.* 1989, 10 (8), 982–1012.
- (207) Geißler, G., Torsten. Entwicklung von Assays zur Untersuchung chemisch-induzierter Trockenstresstoleranz, Martin-Luther-Universität Halle-Wittenberg: Halle (Saale), 2013.
- (208) Geissler, T.; Wessjohann, L. A. *J. Plant Growth Regul.* 2011, 30 (4), 504–511.
- (209) ProBiS server <http://probis.cmm.ki.si/>.
- (210) Konc, J.; Janežič, D. *Bioinformatics* 2010, 26 (9), 1160–1168.
- (211) Konc, J.; Česnik, T.; Konc, J. T.; Penca, M.; Janežič, D. *J. Chem. Inf. Model.* 2012, 52 (2), 604–612.
- (212) Schmitt, S.; Kuhn, D.; Klebe, G. *J. Mol. Biol.* 2002, 323 (2), 387–406.
- (213) Konc, J.; Janežič, D. *J. Chem. Inf. Model.* 2007, 47 (3), 940–944.
- (214) Konc, J.; Janežič, D. *Commun. Math. Comput. Chem. MATCH* 58 (3), 569–590.
- (215) Altschul, S. F.; Gish, W. *Methods Enzymol.* 266, 460–480.
- (216) Altschul, S. F.; Madden, T. L.; Schäffer, A. A.; Zhang, J.; Zhang, Z.; Miller, W.; Lipman, D. J. *Nucleic Acids Res.* 1997, 25 (17), 3389–3402.
- (217) Karlin, S.; Altschul, S. F. *Proc. Natl. Acad. Sci.* 1990, 87 (6), 2264–2268.
- (218) Chen, Y.-M.; Shall, S.; O'farrell, M. *Eur. J. Biochem.* 1994, 224 (1), 135–142.
- (219) Korb, O.; Stütze, T.; Exner, T. E. In *Ant Colony Optimization and Swarm Intelligence*; Dorigo, M., Gambardella, L. M., Birattari, M., Martinoli, A., Poli, R., Stütze, T., Eds.; Lecture Notes in Computer Science; Springer Berlin Heidelberg, 2006; pp 247–258.
- (220) R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2014.
- (221) Labute, P. *Pac. Symp. Biocomput. Pac. Symp. Biocomput.* 1999, 444–455.
- (222) Gao, H.; Williams, C.; Labute, P.; Bajorath, J. *J. Chem. Inf. Comput. Sci.* 1999, 39 (1), 164–168.
- (223) Jurik, A.; Reicherstorfer, R.; Zdrzil, B.; Ecker, G. F. *Mol. Inform.* 2013, 32 (5-6), 415–419.
- (224) Fawcett, T. *Pattern Recognit. Lett.* 2006, 27 (8), 861–874.
- (225) Hanley, J. A.; McNeil, B. J. *Radiology* 1983, 143 (1), 29–36.
- (226) Kabsch, W. *Acta Crystallogr. Sect. A* 1976, 32 (5), 922–923.
- (227) Kabsch, W. *Acta Crystallogr. Sect. A* 1978, 34 (5), 827–828.
- (228) Grant, B. J.; Rodrigues, A. P. C.; ElSawy, K. M.; McCammon, J. A.; Caves, L. S. D. *Bioinforma. Oxf. Engl.* 2006, 22 (21), 2695–2696.
- (229) Punta, M.; Coghill, P. C.; Eberhardt, R. Y.; Mistry, J.; Tate, J.; Bourns, C.; Pang, N.; Forslund, K.; Ceric, G.; Clements, J.; Heger, A.; Holm, L.; Sonnhammer, E. L. L.; Eddy, S. R.; Bateman, A.; Finn, R. D. *Nucleic Acids Res.* 2012, 40 (D1), D290–D301.
- (230) Lovell, S. C.; Davis, I. W.; Arendall, W. B.; de Bakker, P. I. W.; Word, J. M.; Prisant, M. G.; Richardson, J. S.; Richardson, D. C. *Proteins Struct. Funct. Bioinforma.* 2003, 50 (3), 437–450.
- (231) Colovos, C.; Yeates, T. O. *Protein Sci.* 1993, 2 (9), 1511–1519.
- (232) Wiederstein, M.; Sippl, M. J. *Nucleic Acids Res.* 2007, 35 (suppl 2), W407–W410.
- (233) Sippl, M. J. *Proteins Struct. Funct. Bioinforma.* 1993, 17 (4), 355–362.

- (234) Sippl, M. J. *Curr. Opin. Struct. Biol.* 1995, 5 (2), 229–235.
- (235) Bowie, J. U.; Luthy, R.; Eisenberg, D. *Science* 1991, 253 (5016), 164–170.
- (236) Lüthy, R.; Bowie, J. U.; Eisenberg, D. *Nature* 1992, 356 (6364), 83–85.
- (237) Cleveland, W. S. *J. Am. Stat. Assoc.* 1979, 74 (368), 829–836.
- (238) Penning, T. D.; Zhu, G.-D.; Gong, J.; Thomas, S.; Gandhi, V. B.; Liu, X.; Shi, Y.; Klinghofer, V.; Johnson, E. F.; Park, C. H.; Fry, E. H.; Donawho, C. K.; Frost, D. J.; Buchanan, F. G.; Bukofzer, G. T.; Rodriguez, L. E.; Bontcheva-Diaz, V.; Bouska, J. J.; Osterling, D. J.; Olson, A. M.; Marsh, K. C.; Luo, Y.; Giranda, V. L. *J. Med. Chem.* 2010, 53 (8), 3142–3153.
- (239) Halpern SD; Karlawish JT; Berlin JA. *JAMA* 2002, 288 (3), 358–362.
- (240) Field, A. *PsyPAG Q.* 2006, 58, 9–23.
- (241) Ellis, P. D. *The essential guide to effect sizes: statistical power, meta-analysis, and the interpretation of research results*; Cambridge University Press: Cambridge; New York, 2010.
- (242) Cohen, J. *Statistical Power Analysis for the Behavioral Sciences, Second Edition*; Routledge, 1988.
- (243) Triballeau, N.; Acher, F.; Brabet, I.; Pin, J.-P.; Bertrand, H.-O. *J. Med. Chem.* 2005, 48 (7), 2534–2547.
- (244) Schmidt, F. L. *Psychol. Methods* 1996, 1 (2), 115–129.
- (245) Neyman, J.; Pearson, E. *Philos. Trans. R. Soc.* 1933, 231, 289–337.
- (246) Hubbard, R.; Bayarri, M. J. *Am. Stat.* 2003, 57 (3), 171–178.
- (247) *ConfGen, version 2.2*; ConfGen, version 2.2, Schrödinger, LLC, New York, NY, 2010., 2010.
- (248) Hoenig, J. M.; Heisey, D. M. *Am. Stat.* 2001, 55 (1), 19–24.
- (249) Onwuegbuzie, A. J.; Leech, N. L. *Underst. Stat.* 2004, 3 (4), 201–230.
- (250) Becker, O. M.; Dhanoa, D. S.; Marantz, Y.; Chen, D.; Shacham, S.; Cheruku, S.; Heifetz, A.; Mohanty, P.; Fichman, M.; Sharadendu, A.; Nudelman, R.; Kauffman, M.; Noiman, S. *J. Med. Chem.* 2006, 49 (11), 3116–3135.
- (251) Sun, H. *ChemMedChem* 2006, 1 (3), 315–322.
- (252) Song, M.; Clark, M. *J. Chem. Inf. Model.* 2006, 46 (1), 392–400.
- (253) Becker, O. M.; Marantz, Y.; Shacham, S.; Inbal, B.; Heifetz, A.; Kalid, O.; Bar-Haim, S.; Warshaviak, D.; Fichman, M.; Noiman, S. *Proc. Natl. Acad. Sci. U. S. A.* 2004, 101 (31), 11304–11309.
- (254) Zhou, Y.; Peng, H.; Ji, Q.; Zhu, Z.; Yang, C. *Bioorg. Med. Chem. Lett.* 2006, 16 (22), 5878–5882.
- (255) Floquet, N.; Richet, C.; Durand, P.; Bernard, M.; Badet, B.; Badet-Denisot, M.-A. *Bioorg. Med. Chem. Lett.* 2007, 17 (7), 1966–1970.
- (256) Parrill, A. L.; Echols, U.; Nguyen, T.; Pham, T.-C. T.; Hoeglund, A.; Baker, D. L. *Bioorg. Med. Chem.* 2008, 16 (4), 1784–1795.
- (257) White, A. W.; Almassy, R.; Calvert, A. H.; Curtin, N. J.; Griffin, R. J.; Hostomsky, Z.; Maegley, K.; Newell, D. R.; Srinivasan, S.; Golding, B. T. *J. Med. Chem.* 2000, 43 (22), 4084–4097.
- (258) Lehtiö, L.; Jemth, A.-S.; Collins, R.; Loseva, O.; Johansson, A.; Markova, N.; Hammarström, M.; Flores, A.; Holmberg-Schiavone, L.; Weigelt, J.; Helleday, T.; Schüler, H.; Karlberg, T. *J. Med. Chem.* 2009, 52 (9), 3108–3111.

- (259) Miyashiro, J.; Woods, K. W.; Park, C. H.; Liu, X.; Shi, Y.; Johnson, E. F.; Bouska, J. J.; Olson, A. M.; Luo, Y.; Fry, E. H.; Giranda, V. L.; Penning, T. D. *Bioorg. Med. Chem. Lett.* 2009, *19* (15), 4050–4054.
- (260) Gandhi, V. B.; Luo, Y.; Liu, X.; Shi, Y.; Klinghofer, V.; Johnson, E. F.; Park, C.; Giranda, V. L.; Penning, T. D.; Zhu, G.-D. *Bioorg. Med. Chem. Lett.* 2010, *20* (3), 1023–1026.
- (261) Kirby, C. A.; Cheung, A.; Fazal, A.; Shultz, M. D.; Stams, T. *Acta Crystallograph. Sect. F Struct. Biol. Cryst. Commun.* 2012, *68* (Pt 2), 115–118.
- (262) Karlberg, T.; Markova, N.; Johansson, I.; Hammarström, M.; Schütz, P.; Weigelt, J.; Schüler, H. *J. Med. Chem.* 2010, *53* (14), 5352–5355.
- (263) Gunaydin, H.; Gu, Y.; Huang, X. *PloS One* 2012, *7* (3), e33740.
- (264) Naylor, E.; Arredouani, A.; Vasudevan, S. R.; Lewis, A. M.; Parkesh, R.; Mizote, A.; Rosen, D.; Thomas, J. M.; Izumi, M.; Ganesan, A.; Galione, A.; Churchill, G. C. *Nat. Chem. Biol.* 2009, *5* (4), 220–226.
- (265) Zhou, Z.; Wang, Y.; Bryant, S. H. *J. Mol. Graph. Model.* 2010, *28* (8), 714–727.
- (266) Fells, J. I.; Tsukahara, R.; Liu, J.; Tigyi, G.; Parrill, A. L. *J. Mol. Graph. Model.* 2010, *28* (8), 828–833.
- (267) Ijjaali, I.; Petitet, F.; Dubus, E.; Barberan, O.; Michel, A. *Bioorg. Med. Chem.* 2007, *15* (12), 4256–4264.
- (268) Thai, K.-M.; Ecker, G. F. *Bioorg. Med. Chem.* 2007, *16* (7), 4107–4119.
- (269) Gao, H.; Bajorath, J. *Mol. Divers.* 1998, *4* (2), 115–130.
- (270) Yates, F. *Suppl. J. R. Stat. Soc.* 1934, *1* (2), 217.

Curriculum vitae

Persönliche Daten

Name: Peter-Paul Heym
Geburtsdatum: 22.08.1983
Geburtsort: Hoyerswerda
Familienstand: Ledig, 2 Kinder
Nationalität: Deutsch
Vater: Wolfgang Lutz Heym
Mutter: Gabriele Heym

Schulbildung und Studium

1994-2002 Allgemeine Hochschulreife (Abitur), Note 1,9
2002-2009 Studium Bioinformatik
Martin-Luther-Universität Halle-Wittenberg
Abschluss: Diplom, Note 1,6
2009-2015 Promotion an der
Martin-Luther-Universität Halle-Wittenberg
2010-2011 Graduate Certificate in Statistics
University of Sheffield

Beruf

02/2009-09/2014 mit Unterbrechung; wissenschaftlicher Mitarbeiter
am Agrochemischen Institut Piesteritz e.V. (AIP)
01/2012-08/2012 Mitarbeiter bei Startblock24.de
07/2012-09/2012 wissenschaftlicher Mitarbeiter am Institut für
Pflanzenbiochemie Halle (IPB Halle)
seit 10/2014 Einzelunternehmer
Sum Of Squares Peter-Paul Heym
Elternzeit: 11/2009, 06/2010, 05/2011,
sowie 01/2013-06/2013

Publications

Scientific publications

Virtual screening for plant PARP inhibitors – what can be learned from human PARP inhibitors?

P.-P. Heym, W. Brandt, L. A. Wessjohann, and H.-J Niclas

J Cheminform. 2012; 4(Suppl 1): O24.

published online: 2012 May 1. doi: 10.1186/1758-2946-4-S1-O24

In silico Characterization of *At*PARP and Virtual Screening for *At*PARP Inhibitors

P.-P. Heym, T. Geissler, S. Pienkny, W. Brandt, L. A. Wessjohann, and H.-J Niclas
in preparation

Conference talks

Virtual screening for plant PARP inhibitors – what can be learned from human PARP inhibitors?

P.-P. Heym, W. Brandt, L. A. Wessjohann, and H.-J Niclas

7th German Conference on Chemoinformatics.

06.11.-08.11.2011, Goslar, Germany

Poster presentations

A distance-based method for evaluating protein-structures

P.-P. Heym, W. Brandt, L.A. Wessjohann

4. German Conference of Chemoinformatics”

09.11.-11.11.2008, Goslar, Germany

Protein modelling of *Arabidopsis thaliana* L. PARP-1 and pharmacophore design

P.-P. Heym, W. Brandt, L.A. Wessjohann

Molecular Modelling workshop

07.09.-11.09.2009, Erlangen, Germany

PARP inhibitors as agrochemically active substances

T. Geissler, P.P. Heym, L. A. Wessjohann

PARP2010 - 18th International Conference on ADP-ribose metabolism

18.08.-21.08.2010 Zürich, Switzerland

Virtual screening for plant PARP inhibitors to increase tolerance against abiotic stress

P.-P. Heym, W. Brandt, L.A. Wessjohann

9th triennial WATOC congress

17.07.-22.07.2011, Santiago de Compostela, Spain

Virtual screening for plant PARP inhibitors to increase tolerance against abiotic stress

P.-P. Heym, W. Brandt, L.A. Wessjohann

BIOTEC-FORUM - Bioinformatics and Computational Biology

06.12.-07.12.2012, Dresden, Germany

PARP-like proteins and their contribution to plant stress responses

Rissel, D. Heym, P.-P., Peter, K., Thor, K., Brand, W., Peiter, E.

Plant Nutrition 2014 International Conference

10.09.- 12.09.2014 Halle (Saale), Germany

Virtual screening for plant PARP inhibitors to increase tolerance against abiotic stress

Heym P.-P., Geissler T., Berger R., Brandt W., Niclas H.-J., Wessjohann L. A.

Plant Nutrition 2014 International Conference

10.09.- 12.09.2014 Halle (Saale), Germany

Eidesstattliche Erklärung

Ich erkläre an Eides statt, dass ich die Arbeit selbstständig und ohne fremde Hilfe verfasst, keine anderen als die von mir angegebenen Quellen und Hilfsmittel benutzt und die den benutzten Werken wörtlich oder inhaltlich entnommenen Stellen als solche kenntlich gemacht habe.

Halle (Saale), 24.03.2015

Statement under oath

I declare under oath that this thesis is my own work entirely and has been written without any help from other people. I used only the sources mentioned and included all the citations correctly both in word or content.

Halle (Saale), 24.03.2015