1. Abstract

My thesis includes four research projects:

Project 1:	Study of MDM2 binding to p53-analogues: affinity, helicity and applicability to drug design
Project 2:	Allosteric inhibition of HIV-1 integrase probed with normal mode analysis
Project 3:	Structure-based design of small molecules inhibiting HIV integration
Project 4:	New model of CFTR proposes conducting state conformation

1.A. Project 1:

MDM2 is a key regulator of the p53 tumor-suppressor protein. In this project I studied the relative affinity of a p53 N-terminal fragment and its analogs to MDM2 using implicit-solvent MD and MM-GB/SA calculations. Notably: 1) The effect of regulatory phosphorylations at Ser/Thr residues was addressed, suggesting that a balance between favorable electrostatics and desolvation penalties determines the effect of phosphorylation; 2) A helix stabilizing role was proposed for several residues involved in hydrogen bonding and hydrophobic packing; 3) Good correlations between calculated and experimental affinities were obtained for a set of peptidomimetic inhibitors, both alone and in combination with p53 analogues, demonstrating potential applicability to drug design. From the technical aspect, protocol optimization and selection of simulation tools were addressed in detail. To the best of my knowledge this was the first example of MM-GB/SA calculations utilizing a conformational ensemble generated with implicit solvent MD at the time this project was designed and executed. The results presented suggest that this highly efficient variant of classical explicit-solvent MM-GB/SA may be useful for the study of protein-protein interactions and for the design of peptidomimetic drugs. The work was published in the Journal of Chemical Information and Modeling (Kalid and Ben-Tal 2009, publication enclosed).

1.B. Project 2:

HIV-1 Integrase (IN) catalyzes the integration of reverse-transcribed viral DNA into the host genome. It binds the viral long terminal repeats (LTRs) as a dimer and integrates viral DNA into the host genome as a tetramer, following recruitment of the host protein LEDGF/p75. Peptides derived from the LEDGF IN-binding domain (IBD) have recently been found to inhibit integration through allosteric stabilization of an inactive form of the IN tetramer. A similar effect was proposed for the isolated IBD. I modeled the 3D structure of the inactive IN tetramer, unbound as well as in complex with the IBD and LEDGF-derived peptides, and studied the effect of ligand binding on tetramer stability using the Gaussian Network Model of protein dynamics. I carried out an analysis of differences in mean square fluctuations upon binding as well as RMS deviation in inter-residue cross-correlations along the lowest modes of motion. The results of this analysis demonstrate that ligand binding at the LEDGF binding sites has a long range effect on tetramer dynamics, reaching well into the tetramerization interface. In addition, the analysis proposes that the IBD and IBD-derived peptides have a different effect on tetramer dynamics and possibly bind in distinct stoichiometries. Finally, differences in the magnitude of the dynamic effect of different peptides correlate well with differences in their experimentally determined stabilization capacity.

1.C. Project 3:

The protein-protein interaction between the HIV-1 integrase protein (IN) and its host activator LEDGF is a novel, recently elucidated, drug target. In this work I have constructed a specialized structure-based virtual screening procedure and used it to screen for novel inhibitors of the IN-LEDGF interaction that may serve as leads for the development of new anti-HIV therapeutics. Eighty six compounds were selected for *in-vitro* testing, representing a variety of novel chemical scaffolds with the potential of inhibiting IN function at least as potently as the previously reported inhibitors. So far, 45 compounds were tested for direct inhibition of IN activity, in line with the proposed allosteric mechanism described in Project 2, enabling compound activity both in the presence, and absence, of LEDGF. Seventeen of these were found to be more active than previously published inhibitors used as reference. Activities (obtained at 1:1000 enzyme to compound ratio) range from 62% to 100% inhibition, with three compounds showing 100% inhibition, six over 90% and nine compounds over 80% inhibition. In comparison, reference compound activity was under 60%. In addition, several of the less active compounds were found to be very potent in the presence of LEDGF, suggesting that while their allosteric properties are not strong enough, they are able to efficiently displace LEDGF, which is the biologically more relevant function. Subsequent computational analysis suggested that some, but not all, may also be active-site inhibitors and are therefore potentially dual-acting compounds. All previously published compounds contain carboxylic acids or bioisosteric replacements thereof. The active compounds I discovered are all chemically novel, including carboxylic acids, bioisosters and neutral chemotypes. A patent application is in preparation.

1.D. Project 4:

The cystic fibrosis transmembrane conductance regulator (CFTR) is an unusual ABC transporter, functioning as a chloride channel critical for ion and fluid homeostasis in multiple organs. Disruption of CFTR function is associated with the Cystic Fibrosis (CF) disorder, making it an attractive target for CF drug discovery. In addition, CFTR blockers are being developed as potential anti-diarrheals. CFTR drug discovery is hampered by the lack of high resolution structural data and considerable efforts have been invested in modeling the channel structure. Several model structures of CFTR have been published so far, all derived by homology to the bacterial transporter Sav1866. These models mostly agree with experimental data relating to the overall protein structure, but present the channel in an outward-facing conformation that does not agree with the expected conducting state structure. In this work I collaborated with Dr. James Dalton from UPF and Maya Schushan from the Ben-Tal lab to develop a model of CFTR in its "channel-like" conducting state. This was accomplished with a unique modeling approach combining homology modeling and ROSETTA refinement. In contrast to those previously published, the present model is in agreement with expected channel properties such as pore shape and dimensions, relative solvent accessibility of pore residues and experimentally derived distances between amino acids. I have used the model to explore the interaction of anionic open channel blockers with the CFTR pore, revealing a common binding mode at the top of the inner vestibule, dominated by ionic interactions with K95, in agreement with experimental data, as well as interactions with W1145 and R352. The binding site structure and binding mode hypothesis were further validated using a virtual screening enrichment experiment. I have used the *wt* model to derive a model of the mutant F508del CFTR and subjected both models to MD simulations revealing a destabilizing effect of the F508 deletion. This is in accord with conclusions emerging from experimental studies. Modeling and MD simulations alike highlight previously unaddressed salt-bridge interactions that may be important for structure stability. This work has been submitted for publication.

2. Introduction

2.A. Project 1:

The tumor-suppressor protein p53 leads to cell-cycle arrest and apoptosis in response to cellular damage, acting mostly as a transcription factor. While not required for normal growth and development, p53 is critical for the prevention of tumor development. Defects in the p53 pathway are found in most human cancers and p53 knockout mice show a dramatic increase in tumor occurrence[1]. Due to its powerful inhibition of cell growth, p53 function must be tightly regulated in normal human cells. This is accomplished by multiple control mechanisms, the most important of which seem to be at the mature protein level, including regulation of protein stability, cellular localization, posttranslational modifications and conformational changes affecting its ability to bind DNA[1]. One of the central components in the regulation of p53 is the p53-interacting protein MDM2. MDM2 is in itself a transcriptional target of p53 and so establishes a negative feedback loop in which p53 promotes the expression of its own down regulator. MDM2 amplification has been described in certain tumor types, including some sarcomas and brain cancer. In mice, homozygous deletion of MDM2 leads to apoptosis in the embryo and very early embryonic lethality. This effect is countered by a simultaneous knockout of p53. MDM2 interacts with the N-terminal domain of p53 (residues 15–29), containing the major p53 transcription activation domain. This interaction directly interferes with p53's ability to contact transcriptional co-activator proteins. In addition, MDM2 also functions as a ubiquitin ligase for both p53 and itself and so modulates the half-life of p53[1].

The interaction of p53 with MDM2 has been studied extensively. The following section highlights selected studies from recent years, emphasizing the diversity in acquisition methods, discussing interpretability and assessing compatibility between the studies.

One of the first studies, by Picksley et al., tested the effect of alanine mutations on the MDM2 binding capacity of p53(16-25). Using an ELISA binding assay they showed that residues 18-23 (TFSDLW) are all critical for binding[2]. A study by Lin et al. tested the ability of p53 mutants to bind MDM2 using a single concentration immunoprecipitation assay[3]. This study showed that mutations Leu22Gln and Trp23Ser reduced MDM2 binding to 56% and 22%, respectively while the double mutant retained only 2% of the *wt* activity. However, given only single concentration data, the actual effect on binding affinity is not clear.

Bottger et al. described a series of MDM2 binding peptides selected from a phage display library[4]. This analysis was later extended to a synthetic library of peptides derived from the initial discovery[5]. Bottger et al. characterized the ability of these peptides to inhibit p53 binding to MDM2 using three different ELISA formats (peptide-ELISA, MDM2-ELISA and p53-ELISA, according to the immobilized species). A systematic alanine scan of the optimal phage-derived 12-mer peptide revealed that Phe19 and Trp23 are critical for binding, all substitutions leading to complete loss of activity. (The paper also mentions a similar effect on the *wt* p53 but the data is not shown). Leu26 could only be replaced by Ile, Met and Val albeit with 3-5 fold reduction in potency. In the crystal structure of p53 with human MDM2 (hence MDM2) (1YCR[6]) Phe19, Trp23 and Leu26 form the anchoring interactions of p53 to MDM2, explaining the intolerance for substitutions at these positions. With regard to the importance of Leu26, Bottger et al. also measured a significantly lower affinity for p53(18-25), 50-140-fold weaker than p53(16-27)[5]. This is in contrast to the results of Picksley et al., showing increased potency for p53(18-25) over p53(15-30)[2]. The source for this inconsistency is unclear. Additionally, Bottger et al. demonstrated a 2-fold increase in activity for p53(16-27)Leu22Tyr over *wt*. For the phage derived peptide, position 22 (numbering according to *wt* sequence) was shown to prefer all aromatic residues over the *wt* Leucine. In the 1YCR crystal structure, Leu22 is involved in vdW interactions with MDM2 residues Lys94 and His73. Superimposition of the crystal structures of MDM2 in complex with p53(15-29) (1YCR) and with a 9-mer peptide from the Bottger study (1T4F[7]) suggests that the preference for aromatics may be exclusive to the Bottger peptides, which lack Glu17 (the longer analogues have a proline at this position). In the 9-mer complex Try22 is involved in T-shaped aromatic interactions with His73 of MDM2. However, in the *wt* complex Glu17 occupies the same region, making His73 significantly less accessible to interactions with position 22 (Figure 1). Quantitative data could not be extracted from this study due to the variability between the three ELISA assays.

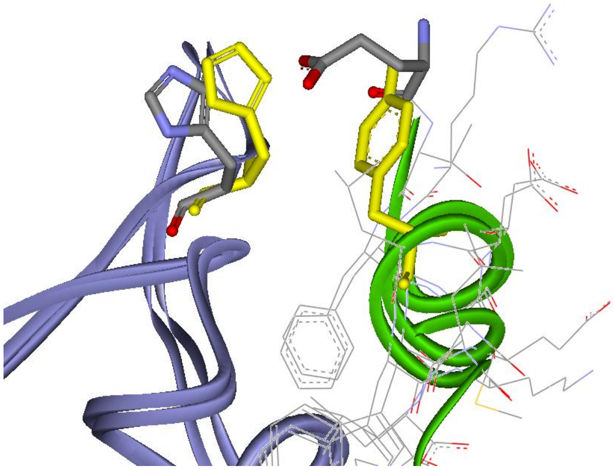


Figure 1. **Preference for Tyr at position 22**: Superimposition of the 1YCR and 1T4F crystal structures. Green ribbon: bound peptide ligand. Blue ribbon: MDM2. In the 1T4F complex of MDM2 with the Bottger 9-mer peptide, Try22 (sticks, yellow) is involved in T-shaped aromatic interactions with His73 (sticks, yellow). However, in the *wt* complex, Glu17 (sticks, colored by element) occupies that same space, making His73 (sticks, colored by element) significantly less accessible to interactions with position 22. Therefore the preference for aromatic interactions may be unique to the Bottger peptides, which all lack Glu17 (the longer analogues have Pro residue at this position).

A study by Lai et al. probed the effects of phosphorylations and peptide length on the p53-MDM2 interaction using a fluorescence anisotropy competition assay[8]. Lai et al. showed that phosphorylation of either Ser15, Ser20 or a dual phosphorylation of both had a minor effect on peptide affinity (Table 1), while phosphorylation of Thr18 alone or in combination with Ser15 led to a ~20-fold reduction in binding. They also showed that the affinity of the 35 residue long peptide p53(1-35) (700nM) is identical to that of the shorter p53(10-29) (750nM), in accordance with the crystal structure of the p53-MDM2 complex (1YCR[6]), which includes only residues 17-29 of p53. Lai et al. were the first to demonstrate the tighter binding of the N-terminal truncated peptide, p53(18-26), having an affinity of 70nM. Further truncation to p53(19-26) caused a 10-fold reduction in affinity, revealing the critical role of Thr18 in p53 binding to MDM2, and a deletion of the critical Leu26 led to an effectively inactive peptide p53(19-25) (Table 1).

Table 1. Collection of p53 derivatives used in the present analysis[a]

	Kd (μM)	*Error(μM)*
Full Length Peptides		
p53(1-35) [b]	0.7	0.1
p53(7-36) [c]	0.121	0.001
p53(12-30)[c]	0.229	0.012
p53(10-29) [b]	0.75	0.1
p53(15-29) [d]	0.575	0.019
PMD13: p53(15-29)F19Nal [d]	0.84	0.06
PMD14: p53(15-29)W23Nal [d]	1.35	0.4
PMD16: p53(15-29)L26F [d]	1.27	0.12
PMD27: p53(15-29)T18S [d]	0.47	0.06
p53(12-30)D21A [c]	0.17	0.007
p53(12-30)W23L [c]	43	2
p53(12-30)L26I [c]	0.209	0.014
p53(12-30)P27S [c]	0.0047	0.0006
Short Peptides		
p53(17-26) [b]	0.05	0.01
p53(18-26) [b]	0.07	0.01
p53(19-26) [b]	0.8	0.1
p53(19-25) [b]	150	50
PMD2: p53(17-26) [d]	0.046	0.007
PMD3: p53(17-26)F19Nal [d]	0.053	0.002
PMD4: p53(17-26)W23Nal [d]	0.93	0.23
PMD22: p53(19-26) [d]	>5	
PMD23: p53(17-26)F19MePhe [d]	1.8	0.2
PMD25: p53(18-26) [d]	0.07	0.009
Phosphorylated Peptides		
p53(10-29)pS15 [b]	1.29	0.15
p53(10-29)pS20 [b]	0.55	0.1
p53(10-29)pS15/pS20 [b]	0.63	0.1

p53(10-29)pT18 [b]	11	1
p53(10-29)pS15/pT18 [b]	14	1
p53(15-29)pS15 [d]	0.27	0.006
p53(15-29)pS20 [d]	0.36	0.006
p53(15-29)pT18 [d]	3.15	0.28
p53(15-29)pS15/pS20 [d]	0.2	0.035
p53(15-29)pS15/pT18 [d]	1.317	0.3
p53(15-29)pS20/pT18 [d]	2.3	0.48

[a] Experimental K_d and measurement error are given for each peptide, as reported in the reference specified in the last column.
[b] Lai et al.[8]
[c] Zondlo et al.[9]
[d] Schon et al.[10]

A single-concentration binding assay by Jabbur et al. demonstrated the importance of the Thr18-Asp21 hydrogen bond network for p53 binding[11]. At the tested concentration, p53(1-52)Thr18Ala, p53(1-52)Thr18Asp and p53(1-52)Asp18Ala retained 65% and 47% and 26% of *wt* activity, respectively, while the non hydrogen-bond disrupting mutation Asp21Asn retained full activity.

A more recent comprehensive study by Schon et al. examined the interaction of MDM2 with a peptide library based on residues 15-29 of p53 using isothermal titration calorimetry (ITC), revealing the effects of peptide length, natural as well as non-natural amino-acid substitutions and phosphorylation at Ser/Thr residues[10]. In this study Phe19 and Trp23 were replaced with the non-natural amino-acids naphthylalanine (Nal) with a limited loss of activity (~1.5-fold and ~2-fold for Phe19 and Trp23, respectively). In the crystal structure, Trp23 makes multiple vdW interactions within the MDM2 binding cleft as well as a hydrogen bond with the carbonyl oxygen of MDM2 Leu54. It is not immediately clear why the loss of a hydrogen bond has such a minor effect on binding. A possible explanation is that it is partially exposed to the solvent. Indole groups are weak hydrogen bond donors, further weakened with increased polarity of the surroundings[12]. While the weak hydrogen bond is lost, the larger Nal group may gain compensating vdW interactions and may also reduce peptide desolvation penalty due to increased hydrophobicity. This study also demonstrated the tighter binding of shorter peptides (Table 1) and the detrimental effect of the Thr18 phosphorylation. In contrast to the results of Bottger et al.[5], Leu26 was shown to tolerate a Phe substitution with a mere 2-fold loss in activity. Surprisingly, an identical binding affinity was measured for p53(17-26) and p53(18-26), demonstrating that the highly conserved Glu17 is not important for binding in the context of p53-analogues terminating at Leu26.

A study by Zondlo et al. provided interesting results for several p53 mutations, measured in a fluorescence polarization assay on the background of p53(12-30)[9]. In this study, Leu26Ile had no effect on binding while the Trp23Leu mutation was detrimental, reducing peptide affinity from 229nM to 45μM. This study was the first to demonstrate that a single mutation to Pro27 can have a significant positive effect on binding, Pro27Ser improving affinity of p53(12-30) by 100-fold to 4.7nM. This result implies that the conserved Pro27 may be reducing the inherent affinity of p53 to MDM2, thereby protecting p53 from excessive down regulation by MDM2. p53(12-30)Pro27Ser was examined by CD and NMR and shown to be significantly more helical than *wt* p53(12-30)[9], which improves the peptide's conformational predisposition towards binding. Interestingly, in this study Asp21Ala had almost no effect on binding, in contrast to the results of others[2,5,11].

It is important to note the effect of phosphorylation on p53 activity. p53 function is regulated by a variety of post-translational modification, the major of which being phosphorylation of Ser/Thr residues by several kinases following cellular stress such as DNA damage[1,13,14] . Phosphorylated Ser15 mediates interaction with the closely related transcriptional co-activating proteins p300 and CBP (CREB Binding Protein) but has little effect on the interactions of p53 with MDM2[8,10,14,15]. This is in agreement with p53 residues Ser15 and Gln16 being disordered in the crystals[6]. Phosphorylation of Ser20 has a pivotal role in regulating the MDM2-mediated turnover of p53 following DNA damage[14] but was also found to have an insignificant effect on p53 binding to MDM2[10]. This is in agreement with crystal structure data showing that Ser20 points out of the binding site, in proximity to p53 residue Lys24, which may gain interaction with the phosphate group. The only phosphorylation disrupting p53 binding to MDM2 is at Thr18[8,10]. It has been argued that the hydrogen bonding network between Thr18 and Asp21 helps initiate the p53 helix, required for binding MDM2 and that phosphorylation of Thr18 would abolish these interactions, leading to a reduction in binding affinity[8]. In addition, it has been suggested that phosphorylation of Thr18 may cause electrostatic repulsion between the modified Thr18 and the proximal Asp21 side chain that could further destabilize the complex[8].

One of the most intriguing results is the tighter binding of short p53 analogues, terminating at Leu26 (Table 1). The NMR findings of Uhrinova et al. may provide an explanation for the higher affinity of these peptides[16]. This study showed that in the unbound form of MDM2, residues Gln18–Gln24 appear to have some order, invading the shallow end of the p53-binding groove which is occupied by p53 residues 27–29 in the crystal structure of the complex (Figure 2). Especially notable is the position of Ile19, which occupies the same space taken by Pro27 of p53 and may compete with Pro27 for interactions with Tyr100 of MDM2[16]. Gln18-Gln24 are suggested to form a flexible lid over the p53 binding-site, which is displaced by p53 upon binding. p53 analogues terminating at Leu26 may possibly coexist with the closed lid, requiring a smaller conformational rearrangement in MDM2 than peptides extending beyond position 26. Uhrinova et al. proposed this theory as an explanation for the tighter binding of short peptides. A recent NMR study by Showalter et al.[17] revealed that the N-terminal lid of MDM2 slowly interconverts between a "closed" state, associated with the p53 binding cleft, and an "open" state which is highly flexible. While apo-MDM2 predominantly populates the "closed" state, binding of p53(17-29) shifts the equilibrium towards the "open" state. In contrast, the binding of nutlin-3, a small molecule inhibitor of p53 binding, does not change the relative prevalence of the two states. Crystal structure data imply that this may also be the case for short peptides, terminating at position 26 or before, as suggested by Uhrinova et al. Figure 2 shows an overlap between apo-MDM2 and co-crystal structures of MDM2 in complex with p53(15-29), a 9-mer analogue of p53(18-26) and nutlin-2 (a small molecule inhibitor closely related to nutlin-3[18]). While residues 27-29 of p53(15-29) overlap with the lid residues, it seems that both small molecule and short peptide may coexist with the closed lid.

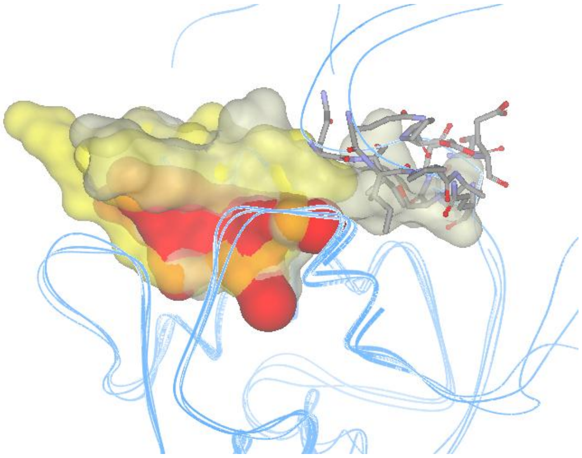


Figure 2. **The "lid hypothesis"**. The 1YCR, 1T4F and 1RV1 complexes of MDM2 with p53, 9-mer peptide and small molecule inhibitor, respectively, superimposed with representative NMR conformations of apo-MDM2 from 1Z1M (blue ribbons). In the apo-structure, the proposed lid residues, Gln18–Gln24 (sticks in element color), penetrate into the shallow end of the p53 binding-site, occupied by residues 27–29 of the p53 peptide in the 1YCR structure. Both 9-mer peptide (yellow surface) and small molecule inhibitor (red surface) occupy roughly the same space and seemingly are able to coexist with the lid in its "closed" state, while p53(15-29) (grey surface) shows significant overlap with the lid residues and probably pushes it away upon binding. This is in accordance with the results of Showalter et al. showing that nutlin-3 does not interrupt the "closed" state of the lid, in contrast to p53(17-29). This also implies that short peptides, terminating before Pro27 may coexist with closed lid, which may explain their higher affinities as described in the text.

Interestingly, in the context of short peptides, deletion of Glu17 was shown not to affect binding (Table 1), despite favorable electrostatic interactions with Lys94 and His73 of MDM2 observed in the 1YCR crystal structure. Strangely, the extent to which deletion of both Thr18 and Glu17 affects binding is somewhat controversial. According to Schon et al., the Kd of p53(19-26) is over 5μM, 10-fold less potent than p53(15-29) and at least 100-fold less potent than p53(17-26)[10]. In contrast, Lai et al. show that p53(19-26) is similar in activity to p53(10-29) and merely 16-fold less potent than p53(17-26)[8] (Table 1).

A better understanding of these complex effects can be achieved by computational techniques, revealing the dynamic behavior of the protein-ligand complexes and the relative contributions of different energetic terms to complex formation.

Theoretical tools have been used in the past to investigate the p53-MDM2 interaction[19-22]. In a pioneering work, Massova and Kollman[19] applied computational alanine scanning based on MM-PB/SA (Molecular Mechanics Poisson-Boltzman / Surface Area, Figure 3) calculations to study the interaction of MDM2 with p53(16-27) in light of the results of Bottger et al.[5]. Although Phe19, Leu22, Trp23 and Leu26 were successfully identified as the key binding determinants, some predictions seem inconsistent with experimental data: Thr18Ala and Asp21Ala were predicted to have no effect on binding, in contrast to mutational and deletion results[2,5,8,10,11]. The positive effect of the Leu22Tyr mutation was predicted to be due to helix-stabilizing interactions between Tyr22 and Glu17 contributing favorably to binding despite a disruption of the Trp23 and Phe19 hydrogen bonds.

$$\begin{tikzcd}
Lig_{sol} \arrow[d, "-\Delta G_{solv}^{lig}"] & + & Prot_{sol} \arrow[r, "\Delta G_{binding}"] \arrow[d, "-\Delta G_{solv}^{prot}"] & Complex_{sol} \arrow[d, "-\Delta G_{solv}^{complex}"] \\
Lig_{vac} & + & Prot_{vac} \arrow[r, "\Delta G_{vac}"] & Complex_{vac}
\end{tikzcd}$$

$$\Delta G_{binding} = \Delta G_{vac} - \Delta G_{solv}^{lig} - \Delta G_{solv}^{prot} + \Delta G_{solv}^{complex}$$

$$\Delta G_{vac} = \Delta H_{vac} - T\Delta S$$

$$H_{vac} \approx E_{vac} = E_{\text{intra}} + E_{elec} + E_{vdW}$$

1)
$$\begin{aligned}\Delta G_{binding} &= \Delta E_{\text{intra}} + \Delta E_{elec} + \Delta E_{vdW} - \\ &\quad - T\Delta S + \Delta G_{solv}^{compelx} - \Delta G_{solv}^{prot} - \Delta G_{solv}^{lig}\end{aligned}$$

Figure 3: Thermodynamic cycle used for calculating binding free energies in MM-GB/SA. $\Delta G_{binding}$ is the free-energy of binding in solution; $\Delta G^{lig}{}_{solv}$, $\Delta G^{prot}{}_{solv}$ and $\Delta G^{complex}{}_{solv}$ are the solvation free energies of the ligand, the protein and the complex, respectively; ΔG_{vac} is the free energy of binding in vacuum; E_{intra}, E_{elec} and E_{vdW} are the force field intramolecular, electrostatic and vdW energies, respectively. Equation 1 is the final MM-GB/SA equation.

Zhong et al. used the same MM-PB/SA methodology to identify critical residues for binding in both p53 and MDM2 and their results for p53 mutations are similar to those of Massova and Kollman[19]. A more recent paper by Lee et al. addressed the issue of p53 phosphorylation, also using MM-GB/SA[22]. One interesting finding from this study is that phosphorylation at Thr18 does not disrupt the helical conformation of p53 in solution as previously suggested. Local anionicity of MDM2 leading to electrostatic repulsion of the phosphate group was suggested to cause the reduced affinity of pThr18[22]. As expected from experiment, phosphorylation at Ser20 was predicted to have a smaller effect than at

Thr18. However, pSer20 and pThr18 were predicted to be 2- and 4- orders of magnitude less active than *wt*, respectively, in disagreement with experimental data (Table 1). In addition, the double phosphopeptide pSer20/Thr18, which has a Kd of 2.3μM[10] was predicted to be completely inactive. Recently, Dastidar et al.[23] published an explicit-solvent MM-GB/SA study with interesting results for the Pro27Ser mutant. In solution, this peptide was modeled in an α-helical conformation, in accordance with experimental results[9]. However, in complex with MDM2, the peptide was modeled once as an α-helix and once in *wt* conformation. In both cases, the mutant was found to be more active than *wt*. While the helical conformation was found to have an enthalpic advantage, the *wt* conformation had an entropic advantage. It is not clear which conformation exists in reality. Interestingly, when the mutant was modeled as α-helical in the complex, Tyr100 from MDM2 flipped during the simulation from the 1YCR *trans* conformation to the *gauche(-)* conformation common to all co-crystal structures with short peptides and small molecules (e.g. 1T4F, 1T4E, 1RV1, 2GV2, 2AXI). Together with the compactness of the helical conformation, this may indicate that the Pro27Ser mutant binds in α-helical conformation coexisting with the binding-site lid, which may contribute to its high affinity.

Each of the previous computational studies covered a small portion of the available empirical data described above. The abundance of untreated data and the controversy between experiment and some of the computational results provide motivation for the present work, which attempts to address a diverse set of p53 modifications including mutations to natural as well as non-natural amino-acids, regulatory phosphorylations and deletions.

2.B. Project 2:

The HIV-1 Integrase (IN) catalyzes integration of the reverse-transcribed viral DNA into the host genome, a procedure essential for viral replication. The integration process comprises two stages[24]: 1) 3’ end processing, in which IN creates the DNA template for integration by removing dinucleotides from the 3’ ends of both ends of the viral DNA Long Terminal Repeat sequences (LTRs) after reverse transcription in the cytoplasm; 2) strand transfer, which is the integration of the viral DNA template into the target host DNA following nuclear import.

IN exists in equilibrium between dimeric, tetrameric, and high-order oligomeric states[25-27]. Dimeric IN binds at each end of the viral DNA during the 3’ end processing in the cytoplasm[28]. Following nuclear import, the two LTR-bound dimers bind the host protein lens epithelium-derived growth factor (LEDGF)/p75 and assemble into a tetrameric form required for strand-transfer[29].

Recently, Hayouka et al. demonstrated that peptides derived from the LEDGF IBD were capable of inhibiting IN activity by shifting its oligomerization state in the cytoplasm from the dimeric form, which binds LTR DNA and catalyzes the 3’ end processing, to a tetrameric form which is incapable of binding LTR DNA and catalyzing this reaction[30]. This prevents initiation of the integration process and inhibits viral replication both *in-*

vitro and *in-vivo*[30,31]. These authors also suggest that a similar mechanism of inhibition may be possible for the isolated IBD, which was also found to inhibit IN activity[30,32]. The main goal of the present study was to investigate the effect of the LEDGF-derived peptides on the IN oligomeric state using computational tools.

Recently, Gupta et al. proposed a model of an IBD-bound IN tetramer[33]. This model, based on small angle X-ray scattering (SAXS) data obtained in the absence of DNA together with X-ray symmetry and hydrodynamic properties, differs significantly from previously published models, which attempted to capture the DNA bound form[29,34-38]. Residues important for concerted integration are separated by excessive distances[33], suggesting that this model may represent an “inactive” non DNA-binding tetramer. This, in turn, may correspond to the tetrameric form stabilized by the LEDGF-derived peptides mentioned above. This also suggests that inhibition of IN by the isolated IBD is at least partly due to stabilization of the inactive tetramer.

Still, the Gupta modeling work, which utilized crystal symmetry to construct a model combining structural element from a number of different X-ray structures, seems to have been preliminary and the final model (which they kindly provided) suffers from critical faults including miss-positioning of the IBD, forming clashes with IN, several missing loop segments and what appears to be manual positional manipulation of one apical N-terminal domain meant to relieve clashes in the initial model (discussed below).

Therefore, I remodeled the LEDGF-bound tetramer based on the initial assembly reported by Gupta but utilizing a more rigorous modeling approach to resolve conflicts and produce a more realistic model.

Next, I used the model structure and normal mode analysis to investigate the effects of peptide binding on IN tetramer dynamics, as well as the effect of the LEDGF IBD, which is also known to stabilize the tetramer[33].

2.C. Project 3:

The critical role of IN in the viral life cycle, as discussed in 2.B., has made it an attractive target for anti-HIV drug development, culminating in the recent approval of Raltegravir, an active-site blocker, for clinical use[39].

However, while pharmaceutical research has mostly focused on active site inhibitors such as Raltegravir (launched in 2007) and its analog Elvitegravir (Phase-III, www.clinicaltrials.gov), the IN mechanism of action suggests additional points of pharmaceutical intervention. Pursuing new ways of intervention is always advisable given the rapid emergence of drug resistant mutations.

One such possible intervention point is the protein-protein interaction between IN and its host activator protein LEDGF/p75. The main determinants of this interaction have recently been elucidated crystallographically[40], providing a possible framework for structure-based drug design.

In this project I set out to design small molecule inhibitors of the IN-LEDGF protein-protein interaction, which could potentially be used as novel anti-HIV therapeutics. The incentive for this work was the recent discovery of IN-inhibitory peptides derived from the IN-binding loops of LEDGF by my collaborator Prof. Assaf Friedler from the Hebrew University in Jerusalem[30,31].

2.D. Project 4:

CFTR is a chloride efflux channel, expressed in the apical membrane of epithelial cells. CFTR is a member of the human subfamily C of the ATP-Binding Cassette (ABC) transporter family, unique in being the only family member known to function as an ion channel. Disruption of CFTR function, which is imperative for maintaining ion and fluid homeostasis, is known to cause Cystic Fibrosis (CF), the most common lethal autosomal recessive disorder in the Caucasian population (http://www.cff.org/AboutCF/).

To date, over 1,400 mutations in the CFTR gene have been identified as disease causing, varying in resulting disease severity[41] (http://www.cff.org/AboutCF/), the most common being a deletion of a single amino-acid, F508, which is found in at least one allele of ~90% of the patients[41]. The ΔF508 mutation nearly abolishes correct cellular processing of CFTR, probably by disrupting inter-domain contacts required for the stability and correct folding of the multi-domain protein[42-44]. Most of the mutant protein is targeted to endoplasmic reticulum-associated degradation (ERAD) and the few mutant channels that make it to the plasma membrane are characterized by a reduced open probability and thermal instability and are rapidly endocytosed and targeted for lysosomal degradation[45].

CFTR architecture is similar to that of other ABC transporters (Figure 4), comprising two membrane spanning domains (MSDs), each linked, through intracellular loops (ICLs) to a nucleotide binding domain (NBD). Uniquely to CFTR, these two repeated motifs are connected by a structurally disordered regulatory domain (R-domain), which must be phosphorylated by protein kinase A (PKA) to enable channel opening [46]. Two models of *wt*-CFTR have recently been reported. Both Serohijos et al.[47] and Mornon et al.[48] reported models of the full length channel based on the crystal structure of the Sav1866 bacterial transporter[49]. Both models are in agreement with a large body of experimental data, demonstrating that Sav1866 is an adequate homology modeling template for CFTR. However, both models present CFTR in the Sav1866 "outward-facing" conformation[49] resulting in a wide barrel-shaped channel structure lacking the architecture proposed by experiment, which includes outer and inner vestibules separated by a narrow region containing the selectivity filter[50-52].

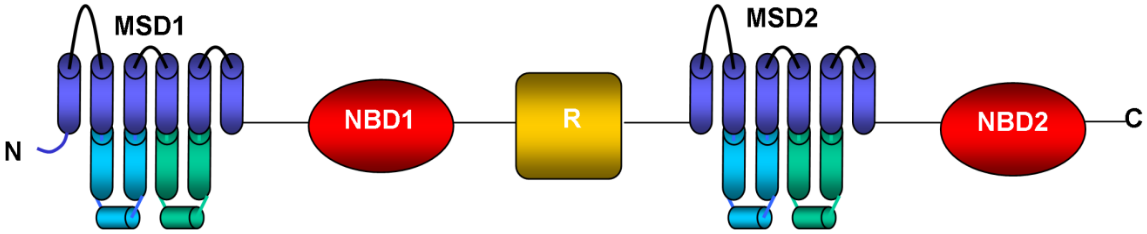


Figure 4: domain arrangement of CFTR.

The objective of the present work was to provide an improved representation of the CFTR active state, by deriving a “channel-like” conformation from the initial outward-facing structure. This was accomplished using a unique modeling protocol guided by a set of experimentally-derived constraints.

The motivation for this modeling project is two-fold. First, such a model could provide further structural insight into the function of CFTR. Second, from a therapeutic perspective, the model may provide an improved structure-based framework for the design of CFTR modulators as potential CF therapeutics[53] as well as channel blockers as potential Antidiarrheals[54].

3. Research Aims

My research objectives in this thesis include the development and application of methodologies for studying pharmaceutically-relevant protein-protein contacts and protein interactions with function-modulating peptides and peptidomimetics, the design of novel protein-protein interaction inhibitors and the modeling of pharmaceutically relevant targets in which protein-protein contacts play a critical functional role.

4. Results & Discussion

4.A. Project 1:

4.A.1. Results:

The following p53-derived peptides were modeled in complex with MDM2 as described in Methods: p53(15-29), p53(15-29)Thr18Ser, p53(15-29)Phe19Nal, p53(15-29)Leu22Tyr, p53(15-29)Trp23Nal, p53(15-29)Trp23Leu, p53(15-29)Leu26Phe, p53(15-29)Pro27Ser; all combinations of single and double-phosphorylations of p53(15-29) at Ser15, Ser20 and Thr18; p53(17-26), p53(17-26)Phe19Nal, p53(17-26)Phe19MePhe, p53(17-26)Trp23Nal, p53(18-26) and p53(19-26); all p53(15-29) alanine mutants.

As discussed in the Introduction and shown in Table 1, there are inconsistencies in the binding affinities reported for several p53 analogues. Since the majority of the peptides studied here were reported by Schon et al.[10], I relied on the binding affinities reported by these authors whenever possible.

As described above, most of the previous computational studies of p53-MDM2 binding used classical MM-P(G)B/SA, employing constant pressure MD simulations of a fully hydrated system[19,20,22,23]. These calculations are highly time-consuming (In my hands ~1 CPU day is required for 100ps of CHARMM[55] simulation on an Intel Xeon E5310 1.60GH with 2GB RAM) and thus are less suitable for studying large datasets. Therefore, I decided to use GB/SA implicit-solvent simulations, which may be less accurate but are

10-fold faster. Implicit-solvent simulations have been used in the past for free energy prediction within the context of Linear Interaction Energy[56] and Thermodynamic Integration[57], but I was not aware of previous reports of MM-GB/SA calculations using ensembles generated with continuum solvent MD at the time this project was designed and executed. Due to the high flexibility of the unbound p53-derived peptides, which may lead to inadequate conformational coverage, I decided to use single-trajectory MM-GB/SA. This method, extracting the conformational ensemble of the unbound protein/ligand from the ensemble of the protein-ligand complex, has been shown to reach better convergence than the multiple trajectory approach relying on independent simulations of the bound and unbound states[19,22,58]. This method neglects the energetic contribution of conformational rearrangements upon binding, and thus inaccuracies may be expected when predicting the relative affinities of peptides with significantly different solution conformations. Calculations were initially performed with CHARMM. However, despite careful optimization of the simulation protocol, stressing the importance of robust temperature control, CHARMM simulations did not yield satisfactory results. Therefore MacroModel[59] Stochastic Dynamics simulations (Methods) were tested as a potential alternative. In my hands, MacroModel was found better suited for this system than CHARMM, and was therefore used here. Details of the CHARMM protocol optimization as well as a comparison of the performance of CHARMM and MacroModel are provided in Appendix A.

On a side note, I first estimated the applicability of single-conformation MM-GB/SA scoring, successfully used by others for ranking congeneric series of compounds[60,61].

Calculations were performed with both CHARMM and MacroModel (Methods), but no correlation was found between experimental affinities and calculated binding energies.

MM-GB/SA binding energies were first calculated for all p53(15-29) non-phosphorylated and single-phosphorylated derivatives with measured binding affinities from the work of Schon et al.[10], obtaining a good correlation with experimental pKd values (R^2=0.77, Figure 5a). The following section provides interpretation for the relative affinities of these peptides and subsequently discusses the effect of double-phosphorylations, mutations studied exclusively by Zondlo et al.[9] and Leu22Tyr, for which binding affinity has not been reported. Prediction results are summarized in Table 2. Calculated free energies of binding are an order of magnitude larger than the experimental ΔG values. This is probably mostly due to the omission of the entropy penalty term (discussed in Appendix A). As noted by Gilson and Zhou, the loss in configurational entropy cancels most of the binding energy, the actual net binding free energy being a small difference between larger numbers[58].

Table 2. MacroModel MM-GB/SA values for p53(15-29) analogues[a]

Peptide	ΔG	ΔG_{GB}	ΔG_{SA}	ΔE_{elec}	ΔE_{vdW}
P53(15-29)	-249.7 (21.3)	1,845.2 (82.6)	-21.0 (1.8)	-1,899.0 (100.1)	-175.0 (23.1)
p53(15-29)F19Nal	-237.7 (19.9)	1,878.7 (51.1)	-9.3 (2.0)	-1,926.2 (64.9)	-180.9 (22.2)
p53(15-29)W23Nal	-246.3 (24.0)	1,921.1 (75.7)	-14.7 (2.4)	-1,958.5 (79.9)	-194.2 (26.9)
p53(15-29)L26F	-210.0 (34.9)	1,770.3 (68.6)	-17.3 (1.8)	-1,801.5 (81.0)	-161.4 (28.3)
p53(15-29)T18S	-257.7 (19.1)	1,849.1 (69.8)	-14.4 (2.3)	-1,913.7 (82.6)	-178.7 (22.4)
p53(15-29)pS15	-293.0 (21.3)	3,050.9 (54.5)	-22.0 (2.0)	-3,156.8 (63.9)	-165.2 (26.7)
p53(15-29)pS20	-279.0 (35.2)	2,737.2 (75.3)	-18.5 (2.1)	-2,817.5 (86.1)	-180.2 (24.2)
p53(15-29)pT18	-217.9 (15.6)	2,542.6 (46.1)	-19.8 (2.0)	-2,560.1 (53.5)	-180.7 (21.8)
p53(15-29)L22Y	-262.9 (18.2)	1,886.5 (61.2)	-16.2 (1.8)	-1,919.2 (70.5)	-214.0 (21.8)
p53(15-29)W23L	-178.0 (17.3)	1,639.3 (57.3)	-18.8 (1.6)	-1,651.6 (61.1)	-146.9 (19.7)
p53(15-29)D21A	-234.6 (21.3)	1,337.6 (60.8)	-19.3 (1.7)	-1,357.0 (75.8)	-195.9 (22.3)
p53(15-29)pT18/S20	-265.1 (20.8)	3,550.1 (83.7)	-19.3 (1.4)	-3,587.4 (91.1)	-208.5 (17.4)
p53(15-29)pS15/T18	-292.5 (22.5)	3,299.2 (71.9)	-20.3 (2.0)	-3,366.0 (84.8)	-205.5 (21.9)
p53(15-29)pS15/pS20	-361.8 (19.2)	3,952.9 (65.5)	-14.8 (2.1)	-4,125.1 (75.4)	-174.8 (22.8)
p53(15-29)E17A	-153.6 (17.3)	1,062.5 (66.0)	-15.5 (2.2)	-1,026.2 (69.5)	-174.4 (21.0)
p53(15-29)T18A	-221.3 (19.2)	1,869.4 (62.0)	-17.6 (1.9)	-1,904.0 (76.0)	-169.2 (20.6)
p53(15-29)F19A	-218.1 (22.0)	1,764.4 (83.4)	-16.6 (1.7)	-1,803.3 (100.7)	-162.5 (23.4)
p53(15-29)L22A	-269.9 (20.3)	1,941.5 (82.9)	-17.0 (1.9)	-2,017.3 (100.0)	-177.1 (21.8)
p53(15-29)W23A	-171.8 (21.0)	1,712.5 (83.4)	-17.1 (1.7)	-1,741.9 (102.6)	-125.4 (23.2)
p53(15-29)K24A	-265.5 (30.0)	2,193.2 (64.9)	-16.7 (1.7)	-2,266.8 (91.3)	-175.1 (24.7)
p53(15-29)L25A	-216.5 (19.8)	1,863.3 (55.4)	-19.2 (2.0)	-1,902.1 (67.9)	-158.5 (20.7)
p53(15-29)L26A	-210.6 (20.8)	1,897.3 (82.6)	-15.7 (1.7)	-1,942.9 (100.0)	-149.3 (21.9)
p53(15-29)P27S *wt-conf*	-245.4 (117.6)	1,568.1 (99.3)	-15.8 (4.4)	-1,573.2 (135.3)	-224.6 (84.9)
p53(15-29)P27S *helical*	-160.1 (21.3)	1,413.6 (153.7)	-13.6 (2.3)	-1,395.3 (154.7)	-164.8 (34.1)
p53(15-29)E28A	-212.2 (30.4)	1,146.4 (183.3)	-20.0 (2.1)	-1,132.2 (207.0)	-206.4 (16.7)
p53(15-29)N29A	-195.8 (19.9)	1,853.8 (101.8)	-16.6 (2.1)	-1,881.5 (115.3)	-151.4 (18.9)

[a] A breakdown of the calculated binding energy (ΔG) into individual components is given for each peptide. ΔG_{GB}, ΔG_{SA}: electrostatic and nonpolar components of the free energy of solvation, respectively. ΔE_{elec}, ΔE_{vdW}: electrostatic and vdW components of the binding enthalpy (See Figure 3 for details). Standard deviation is given in parentheses. p53(15-29)Pro27Ser was modeled in both *wt* and helical conformation.

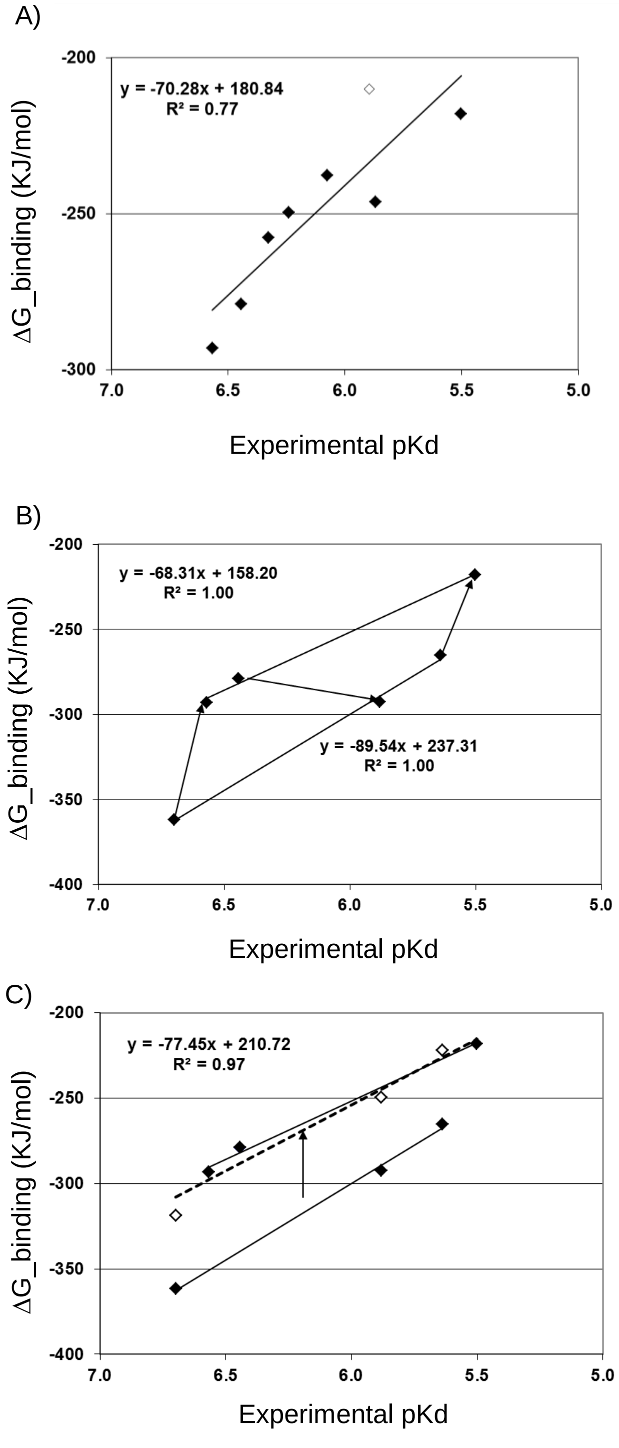


Figure 5. A) Correlation between calculated ΔG and experimental pKd values for a) eight p53(15-29) analogues (first eight in Table 1). Hollow diamond: p53(15-29)Leu26Phe B) Phosphopeptides. Top regression line: single-phosphopeptides. Bottom regression line: double-phosphopeptides. There is a

perfect correlation within each individual series; however the two regression lines are parallel displaced. C). Displacing the double-phosphopeptide series by a constant value equal to the difference between the averages of the two series (hollow diamonds) results in a near perfect correlation with $R^2=0.97$ (dashed line).

p53(15-29)Trp23Nal was predicted to be as active as *wt* p53(15-29). Although the actual predicted affinity is slightly lower, ΔG distributions were compared using student's t-test and were found to be indistinguishable ($p>0.1$). While the reported K_d for this peptide is 1.35μM, the measurement error is 0.4μM, the largest among the peptides tested by Schon et al.[10]. The correlation obtained with MM-GB/SA may be further improved if the lower bound value is used instead of the reported average. Interestingly, in addition to gaining vdW contacts due to the larger Nal group, this peptide also seems to gain electrostatic interactions despite the loss of the Trp23 hydrogen bond. This may indicate that p53(15-29) is capable of compensating for the loss of a hydrogen bond by gaining additional interactions, in contrast to its shorter analogue p53(17-26), discussed below.

p53(15-29)Phe19Nal was predicted to be slightly less active than *wt*, as expected from experiment.

Thr18Ser was predicted to be slightly more active than *wt*, in accordance with experimental results. This seems to be due to improved electrostatic and vdW interactions. During the MM-GB/SA MD simulations, Thr18 did not maintain its hydrogen bond with the Asp21 side-chain. In contrast, Ser18 maintained a stable hydrogen bond with the Asp21 side-chain and also gained additional transient hydrogen bonds with Ser20 and the Asp21 backbone.

MM-GB/SA predicted Leu26Phe to be less active than expected. In this case, the deviation from the regression line is larger than for the other variants included in the correlation (Figure 5a). This problem was addressed by inspecting the distribution of individual MM-GB/SA ΔG values calculated at each sampled MD frame; the final MM-GB/SA energy is an average of these values. Surprisingly, the distribution of ΔG values for Leu26Phe was roughly two-lobed instead of normal as for the other peptides (Figure 6), which may indicate that the calculation has not converged. Indeed, RMSD analysis reveals that the Leu26Phe simulation was significantly less stable than *wt*. It is possible that the initial conformation used for modeling this mutant in complex with MDM2 is not optimal, but this issue was not pursued further. Exclusion of this mutant from the correlation increases R^2 from 0.77 to 0.84.

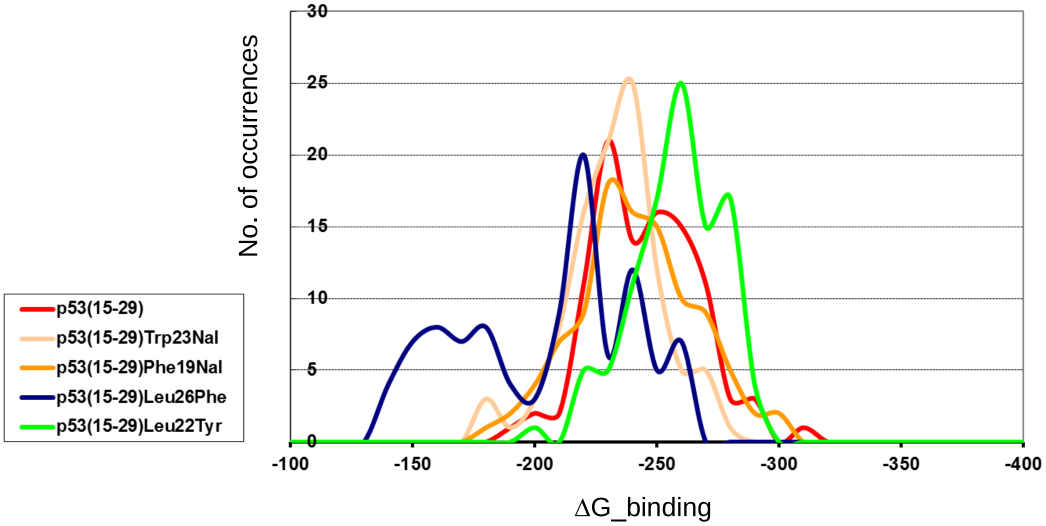


Figure 6. Distribution of calculated ΔG values over the ensemble of conformations that were sampled during the MD simulations. The reported ΔG is the average of these values. ΔG distribution is expected to be roughly normal, which is indeed the case for all the tested peptides (five representative distributions are shown). However, this is not the case for the Ley26Phe peptide (dark blue). This may indicate that the calculation for this mutation has not converged and that the system requires more time to conformationally adjust to the mutation.

Interesting results were obtained for the three single-phosphorylated peptides. The binding affinities predicted for these peptides correlate well with experimental affinities providing a good rank-order both within this group (R^2=1, Figure 5b) and relative to non-phosphorylated peptides (Figure 5a). This supports the credibility of the current computational interpretation which differs significantly from that of Lee et al., described above[22]. According to the current MM-GB/SA calculation, phosphorylations modulate p53 affinity through interplay between two opposing factors: favorable electrostatic interactions and desolvation. For all three peptides there is a significant electrostatic gain. It is largest for pSer15, due to favorable interactions with MDM2 residue Lys94 and the p53 N-terminus (Figure 7a). A smaller gain is predicted for pSer20 interacting with p53 residue Lys24 (Figure 7b). The smallest increase in ΔE_{elec} is predicted for pThr18 which during simulation becomes caged within a network of hydrogen bonds with p53 backbone amides and Gln16 (Figure 7c). These interactions are counterbalanced by a significant increase in the polar desolvation term ΔG_{GB}, providing the correct rank ordering. There is no evidence of electrostatic repulsion due to the presence of an anionic patch as suggested by Lee et al.[22].

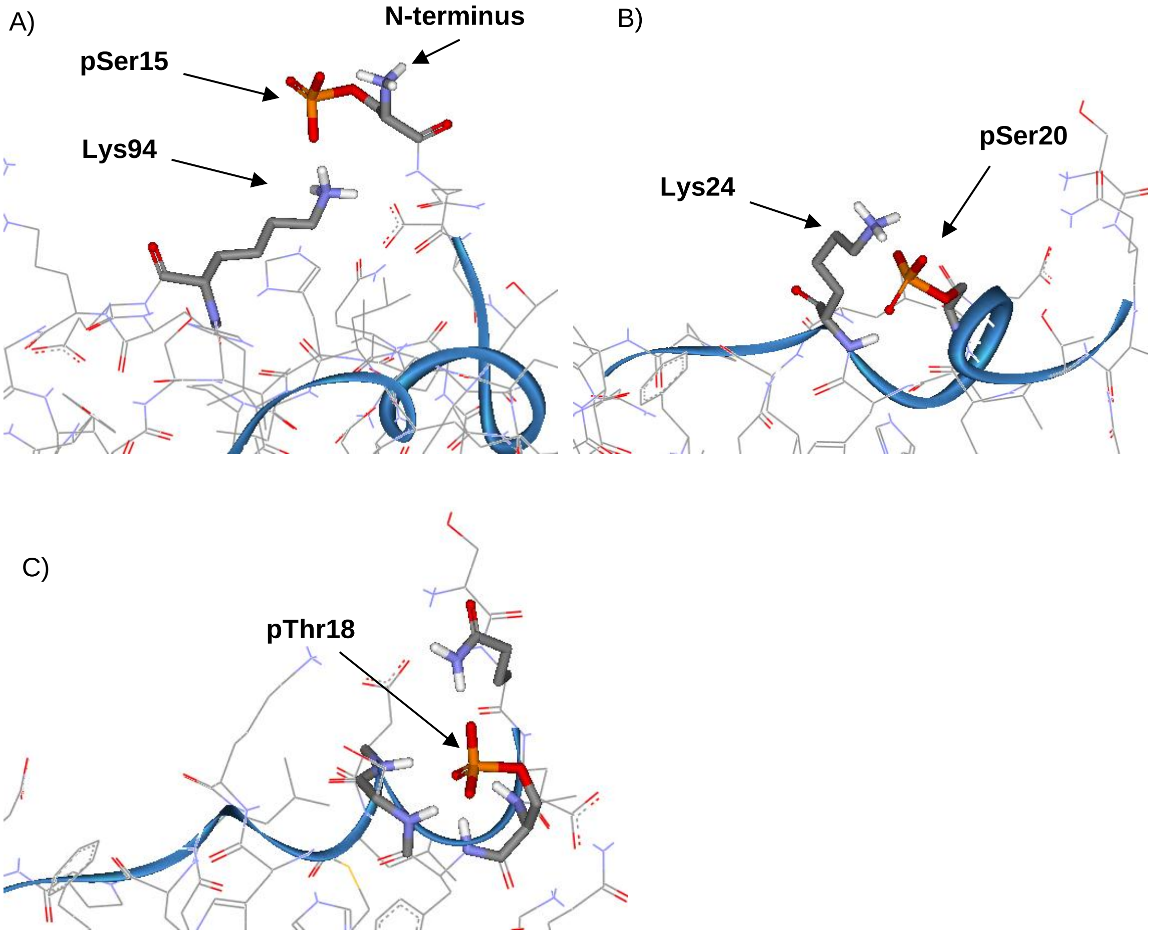


Figure 7. Interactions of phosphate groups observed during the MD simulation within the time frame used for MM-GB/SA sampling. A) pS15 is involved in two salt bridges, with the peptide N-terminus and with MDM2's Lys94. B) pS20 is engaged in a salt bridge with p53 Lys24. C) pT18 is enclosed in a cage of five hydrogen bonds with p53 backbone amides and p53 Gln16 side chain

It has been suggested previously that the lower activity of pThr18 is due to reduced p53 helicity following the loss of the Thr18-Asp21 hydrogen bond[9]. This idea challenges the validity of the current MM-GB/SA interpretation, which provides a good correlation while ignoring differences in the unbound conformational ensemble. Therefore, I performed complementary MD simulations of the unbound *wt* and phosphorylated peptides (Methods) to address the effect of Thr18 phosphorylation on p53 helicity. In accordance with the results of Lee et al.[22] no change in helical stability was observed upon phosphorylation (Figure 8), lending further credibility to my MM-GB/SA results.

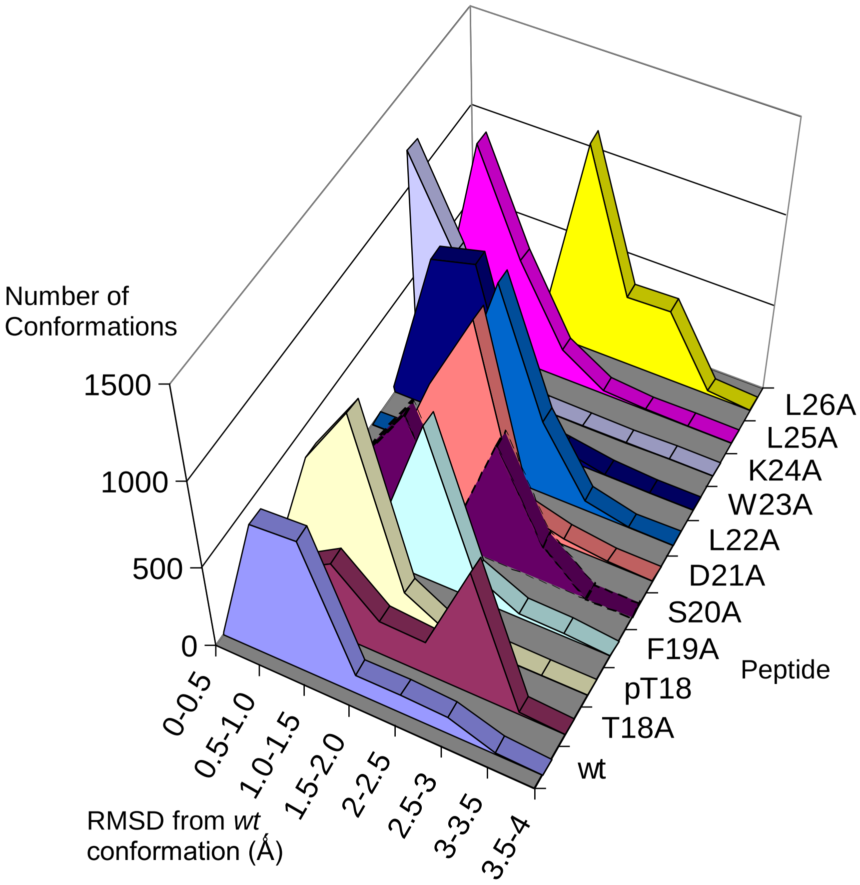


Figure 8. **Stability of bound conformations in solution**. Stability of the p53 helical conformation in solution calculated for *wt* and mutant peptides over a 100ns MD simulation. Stability is represented as the

distribution of p53(19-23) backbone RMSD relative to the *wt* bound conformation. Significant disturbance is seen upon disruption of the Thr18-Asp21-Ser20 hydrogen bond network and upon removal of either Phe19 or Leu26 which pack against Trp23 as well as Leu22. The strongest helix disturbing mutations are Leu26Ala and Thr18Ala. For each p53(15-29) mutant (Y-axis), the number of peptide conformations (Z-axis) is plotted against ranges of RMS deviation from *wt* (Methods).

The results for double-phosphorylated peptides are also very exciting. As for the single-phosphorylated peptides, the correlation between calculated free energies and experimental binding affinities within this group was very good (R^2=1, Figure 4b), ΔG being governed by a balance between favorable electrostatics and desolvation. MM-GB/SA binding energies of the double-phosphopeptides seem to deviate by a constant factor from those of the single-phosphopeptides, as apparent from the two parallel displaced regression lines (Figure 4b). Adding a constant factor to the MM-GB/SA energies of the double-phosphopeptides, equal to the difference between group averages, leads to a perfectly correlated set of binding energies (Figure 4c). Displacement by a constant factor may indicate an overestimation of either ΔE_{elec} or ΔG_{GB} due to the presence of excess charge (double-phosphopeptides have a formal charge of -6 compared with -4 for the single-phosphopeptides and a maximum of -2 for the rest of the peptides analyzed). This may be due to limitations of force field, solvation model or cutoff value used for calculating the electrostatic interactions (Methods) and should be investigated further. Surprisingly, for this set of peptides, implicit-solvent MM-GB/SA appears to be more accurate than the explicit-solvent ns-scale MM-GB/SA of Lee et al., which obtained a positive ΔG value for pThr18/pSer20[22].

In contrast to the mutations discussed above, Trp23Leu, Asp21Ala and Pro27Ser have been tested exclusively by Zondlo et al.[9], on the background of p53(12-30). This p53 segment was reported to have an affinity of 0.229μM, compared with 0.575μM measured for p53(15-29) by Schon et al[10] (Table 1). Due to the lack of a common reference frame, these peptides were excluded from the correlation described above.

As expected, the Trp23Leu mutation was predicted to be detrimental to binding due to a significant loss of protein-ligand interactions.

Asp21Ala was predicted to be less potent than *wt* due to concurrent loss in electrostatic energy and improvement in the solvation energy of the complex. The calculated ΔΔG translates to a ~2-fold activity loss relative to *wt* (Methods), in contrast to a small activity gain reported by Zondlo et al.[9]. However, an activity loss is in agreement with others experimental studies[2,5,11]. Based on crystal structure data, the Asp21-Thr18 hydrogen bond has been suggested to be important for p53 helix initiation[6]. To test this hypothesis, simulations of the unbound mutant peptide in solution were performed, as for pThr18. In accordance with the suggestion above, the mutant peptide was found be significantly less helical in solution than *wt* (Figure 8). This indicates that Asp21 is indeed important for p53 binding, contrary to the results of Zondlo et al., and that p53(15-29)Asp21Ala may be significantly less active that predicted by MM-GB/SA.

Pro27Ser, found by Zondlo et al. to be significantly more α-helical in solution than *wt* p53[9], was modeled twice: once in *wt* conformation, and once in helical conformation (as in Dastidar et al.[23]). In both cases the prediction did not match the experimental data. This was further addressed following the modeling of p53 analogues terminating at Leu26, and is discussed below.

Finally, the predicted effect of Leu22Tyr correlates well with the experimental 2-fold improvement over *wt* reported by Bottger et al.[5]. Specifically, by translating MM-GB/SA energies into predicted pK_d values (Methods) a factor of 1.5 over *wt* affinity is obtained. Relative to Leu22, Tyr22 shows increased vdW and electrostatic interactions. The Tyr22 phenol hydroxyl hydrogen bonds to Glu17 and MDM2 Lys94 and its phenyl ring stacks against the alkyl side-chain of Lys94 (Figure 9), gaining additional vdW interactions compared to the *wt* Leu22. This positive contribution to binding is attenuated by increased ΔG_{GB}. The lack of interaction with MDM2 residue His73 suggests that the aromatic preference may not be as relevant for *wt* p53 as it is for the peptides of Bottger et al. which lack Glu17, as proposed in the Introduction. In contrast to the results of Massova et al.[19], the present calculation did not find a destabilizing effect on p53 hydrogen bonds nor an increased conformational strain in the complex.

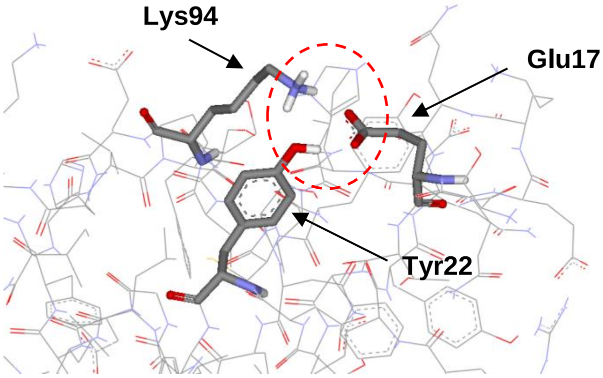


Figure 9. Interactions between Tyr22, Glu17 and MDM2 Lys94 observed during MD simulations within the time frame used for MM-GB/SA sampling. A hydrogen bond network is formed between Try22 phenol hydroxyl, Glu17 and MDM2 Lys94. In addition, the Tyr22 phenyl ring packs against the Lys94 side-chain, gaining additional vdW interactions compared to the *wt* Leu22.

Subsequently, p53 residues 17-29 were consecutively mutated to alanine. Experimental affinities have not been reported for these mutants, except for Asp21Ala discussed above. To facilitate easier comparison between predicted affinities, MM-GB/SA energies were translated into pKd values (Methods). Helical stability was also assessed for each of the mutants (Methods).

p53(15-29)Thr18Ala was found to be significantly less potent than *wt* p53[2,11]. In my calculation, it is predicted to have reduced affinity of ~2μM, mostly due to a larger desolvation penalty than *wt* (Table 2). In addition, its helical stability in solution was strongly diminished (Figure 8), suggesting further reduced binding affinity. This result is in agreement with the helix stabilizing role previously suggested for Thr18[6]. Throughout the simulation of the unbound *wt* peptide, a stable hydrogen bond was observed between the side-chains of Thr18 and Asp17 and an additional, transient, hydrogen bond was found between Thr18 and Ser20. Interestingly, during the MD simulation of the *wt* complex, these hydrogen bonds were mostly lost. This result may indicate that the major role of the Thr18-Asp21 hydrogen bond network is to stabilize the helical conformation in solution prior to binding.

Phe19Ala was predicted to be less active than *wt* due to a loss in ΔE_{vdW} and ΔE_{elec}, compensated for, in part, by a reduction in ΔG_{GB}. However, the effect is not detrimental as expected from experiment (Bottger et al.[5] describe a complete loss of activity, results not shown). In the p53 bound conformation, Phe19 packs against Trp23 and may stabilize the p53 amphipathic helix, which is required for binding. Potential destabilization by Phe19Ala may be the source for the reported loss of activity. Indeed, Phe19Ala was found to be significantly less helical than *wt* p53 (Figure 8).

The effect of the Ser20Ala mutation could not be predicted due to rapid ligand drifting during MD, leading to a complete loss of the Trp23 and Leu26 anchoring interactions. The reasons for this conformational instability, which was not seen in the other simulations, are unclear at this point.

The Leu22Ala mutation was shown by Picksley et al. to be detrimental to the binding of p53(16-25)[2]. In addition, Bottger et al. demonstrated a significant loss of activity for the 12-mer following this mutation[5]. In contrast, MM-GB/SA predicted a slightly improved affinity, mostly due to an increase in the electrostatic component. While MM-GB/SA failed to perceive the effect of this mutation, Leu22Ala was found detrimental to peptide helicity (Figure 8), in accordance with a role in helix formation proposed by others[62].

The Trp23Ala mutant was predicted to be detrimental due to reduced ΔE_{elec} and ΔE_{vdW}, in accordance with experimental results (Table 2). Interestingly, this mutation did not lead to helical instability (Figure 8).

No loss of activity was predicted for the Lys24Ala mutant, in accordance with the Bottger results, revealing a wide tolerance at this position[5].

Leu25Ala was predicted to reduce binding ~5-fold, mostly due to reduced vdW interactions and increased desolvation penalty, contrary to the results of Massova et al.[19]. Interestingly, the temperature factor of the Leu25 side chain in 1YCR is identical to that of the anchoring residue Leu26, suggesting a role in binding. While Bottger et al. demonstrated a wide tolerance at position 25 of the 12-mer[5], the relevance of this result to *wt* p53 is not clear since the amino-acid sequence of the 12-mer is adapted for a glycine at this position.

Leu26Ala was predicted to reduce peptide affinity to ~3μM, mostly due to a higher ΔE_{vdW}. As for Phe19, the effect predicted by MM-GB/SA was not as detrimental as suggested by Bottger et al.[5]. However, Leu26 was also found critical for helical stability (Figure 8) which may bridge the gap.

Glu17Ala, Glu28Ala and Asn29Ala were all found to be less active than *wt*. However, these results are probably unreliable. Despite the seemingly favorable interactions observed in 1YCR, Glu28 and Asn29 are only partially resolved in the crystal[6] and Glu17 has B-factor values as high as those of the former two, indicating a weak contribution to binding of all three residues. The current simulations probably overestimate their contribution.

Following the analysis of p53(15-29) analogues, I addressed the binding of p53 analogues terminating at Leu26 as well as peptidomimetic inhibitors of the p53-MDM2 interaction. The set of peptidomimetics, added to test the applicability of the current MM-GB/SA protocol to drug-design, constituted four helical 8-mer mimetics reported by Garcia-Echeverria et al.[63] (compounds 3, 4, 5 and 8 from that paper) and five beta-hairpins reported by Fasan et al.[64] (compounds 62, 64, 67, 77a and 78a from that paper). Compounds from both papers were selected to span a range of activities (Methods).

Unexpectedly, when the peptides and peptidomimetics were modeled in complex with the structure of MDM2 from 1YCR, no correlation was observed between calculated and experimental affinities. However, good correlations were obtained upon remodeling the protein-ligand complexes using the structure of MDM2 from 1T4F (R^2=0.86 for 8-mer mimetics, Figure 10a; R^2=0.93 for the beta-hairpins; R^2=0.89 for the beta-hairpins combined with peptides terminating at Leu26, Figure 10b). The transition from 1YCR to 1T4F is further discussed in Appendix A. Interestingly, the short p53 analogues, 8-mers and beta-hairpins occupy a similar portion of the MDM2 binding-site, all terminating at a position equivalent to Leu26. Thus, 1T4F may be a better template than 1YCR for assessing the binding of peptides terminating at Leu26 and compounds of similar size. As discussed in Appendix A, the reasons for this remain unclear at this point.

The significant correlations obtained for peptidomimetics alone and in combination with p53 analogues suggest that implicit-solvent MM-GB/SA may support the design of diverse peptidomimetic inhibitors of the p53-MDM2 interaction. While application to small molecule inhibitors is outside the scope of this work, there is no reason for the present methodology not to be applicable in that case as well.

The MM-GB/SA results for p53 analogues obtained with 1T4F-based MM-GB/SA are summarized in Table 3. An interesting result was obtained for p53(17-26)Trp23Nal. The experimental activity loss associated with the Trp23Nal mutation is much more pronounced for p53(17-26) than for p53(15-29) (Table 1). Correspondingly, my calculations showed a much greater reduction in ΔE_{elec} in this case, suggesting that the Trp23 hydrogen bond is more critical for shorter peptides. This may indicate that longer peptides are able to compensate for the loss of this specific interaction by gaining others.

As discussed in the Introduction, p53(15-29)Pro27Ser adopts a helical conformation in solution and was suggested to be helical when bound to MDM2. In the helical conformation, this peptide is similar in size to p53 analogues terminating at Leu26 (size similarity was evaluated by superimposing the mutant peptide in helical conformation onto p53(17-26) in *wt* conformation). Therefore, 1T4F-based MM-GB/SA was tested as an alternative to the 1YCR-based predictions discussed above, which were unsuccessful. Subsequent addition of p53(15-29)Pro27Ser to the set of short p53 analogues and beta-hairpin mimetics maintained the correlation quality (R^2=0.89, Figure 10b). This result indicates that p53(15-29)Pro27Ser may bind MDM2 similarly to peptides terminating at Leu26, possibly without displacing the binding-site lid, which may contribute to its high affinity. This result is in agreement with the simulations of Dastidar et al.[23], showing that p53(15-29)Pro27Ser causes Tyr100 in MDM2 to shift into a 1T4F-like conformation.

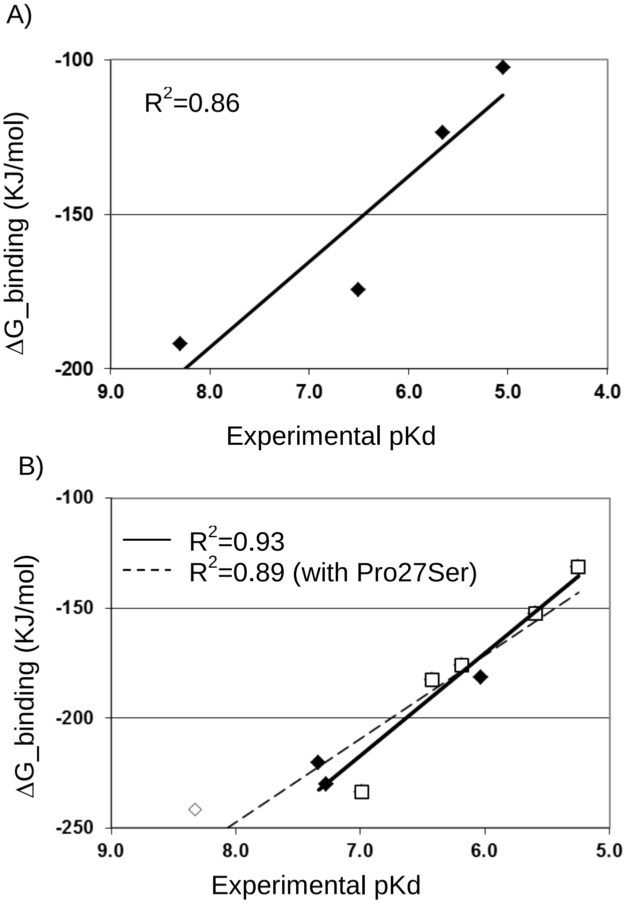


Figure 10. Correlation between calculated ΔG and experimental pKd values for the short peptides and peptidomimetics. A) Correlation obtained for 8-mer peptidomimetics. B) Correlation obtained for beta-

hairpin peptidomimetics in combination with p53(17-26) analogues. Hollow squares: beta-hairpin mimetics. Hollow diamond: p53(15-29)Pro27Ser in helical conformation.

Table 3. MacroModel MM-GB/SA values for short p53 analogues[a]

Peptide	ΔG	ΔG_{GB}	ΔG_{SA}	ΔE_{elec}	ΔE_{vdW}
p53(17-26)	-112.2 (17.7)	932.3 (56.5)	-12.2 (1.9)	-908.6 (66.7)	-123.8 (20.0)
p53(17-26)F19Nal	-122.1 (14.0)	994.8 (37.8)	-9.0 (1.7)	-969.8 (57.8)	-138.1 (17.5)
p53(17-26)W23Nal	-133.7 (12.2)	1,094.9 (44.1)	-7.6 (1.6)	-1,096.2 (50.9)	-124.7 (18.5)
p53(17-26)F19MePhe	-127.2 (15.3)	1,004.5 (34.2)	-10.0 (2.4)	-1,031.3 (45.6)	-90.5 (16.4)
p53(15-29)Pro27Ser-helical	-241.5 (15.9)	1,692.0 (74.8)	-18.1 (2.1)	-1,750.3 (80.7)	-165.1 (20.0)

[a] A breakdown of the calculated binding energy (ΔG) into individual components is given for each peptide. ΔG_{GB}, ΔG_{SA}: electrostatic and nonpolar components of the free energy of solvation, respectively. ΔE_{elec}, ΔE_{vdW}: electrostatic and vdW components of the binding enthalpy (See Figure 3 for details). Standard deviation is given in parentheses. p53(15-29)Pro27Ser was modeled in a helical conformation.

4.A.2. Discussion

Implicit solvent MM-GB/SA was successfully used to predict the relative binding affinities of p53 analogues to MDM2, including natural as well as non-natural amino-acid substitutions, regulatory phosphorylations and peptide truncations. MM-GB/SA energy terms were used in conjunction with structural data generated by MD to provide interpretation of experimental results.

Several conclusions can be drawn within the accuracy limits imposed by the absence of amino- and carboxy-termini of MDM2 and the implicit solvent model: 1) The hydrogen-bond between Thr18 and Asp17 is probably important for stabilizing the helicity of p53 in solution, in preparation for binding. Upon binding, the peptide conformation is stabilized by MDM2 and the hydrogen-bond is lost. This was revealed by a combination of MM-GB/SA and simulations of unbound *wt* and mutant peptides; 2) Phe19, Leu22 and Leu26 stabilize the helical conformation of p53 in solution; 3) Phosphorylation at different positions may regulate p53 MDM2 interaction through a balance between favorable electrostatic interactions and desolvation penalty; 4) The Trp23 hydrogen-bond seems more critical for the binding affinity of short peptides than for long peptides, potentially due to compensating interactions in the long peptides.

From the computational aspect: 1) Implicit solvent MM-GB/SA, which is at least 10-fold more efficient than explicit solvent free-energy prediction methods, is applicable to the analysis of protein-protein interactions. One limitation of the method is that protein-protein interactions take place at shallow, relatively solvent exposed binding-sites, and the absence of water molecules may lead to greater structural drifts during molecular-dynamics. This may limit the length of the molecular-dynamics simulations and therefore the extent of conformational sampling, potentially effecting prediction accuracy; 2) Implicit solvent MM-GB/SA may also be applicable as a predictive tool in the design of protein-protein interaction inhibitors. This was exemplified through correlations obtained for two sets of peptidomimetic inhibitors, both individually and in combination with p53 derivatives; 3) The present results indicate that in the case of shallow solvent-exposed relatively hydrophobic binding-sites which pose a challenge for implicit solvent simulations, MacroModel Stochastic Dynamics with GB/SA and OPLS-2005 may perform better than CHARMM Nose-Hoover dynamics with GB/SW and the CHARMM Momany and Rone force field.

4.B. Project 2:

4.B.1. Results

4.B.1.A. Modeling the non DNA-binding tetramer

An initial model of the inactive IN tetramer was assembled from the crystal structures of HIV-1 IN NTD-CCD (1KY6[37]), HIV-1 IN CCD-CTD (1EX4[65]), HIV-2 IN NTD-CCD in complex with LEDGF (3F9K[36]) and HIV-1 IN CCD in complex with LEDGF (2B4J[40]) according to the general symmetry guidelines provided by Gupta et al.[33] as described in Methods. The assembly was validated using the Gupta model which the authors kindly provided. This initial modeling resulted in severe backbone clashes between the two proximal N-terminal domains (NTDs) from monomers A and C, forming a part of the tetramer interface (Figure 11).

In order to resolve these clashes rigorously, I designed a hybrid modeling protocol in which the limited conformational sampling provided by homology is augmented through integration of protein-protein docking and ROSETTA refinement.

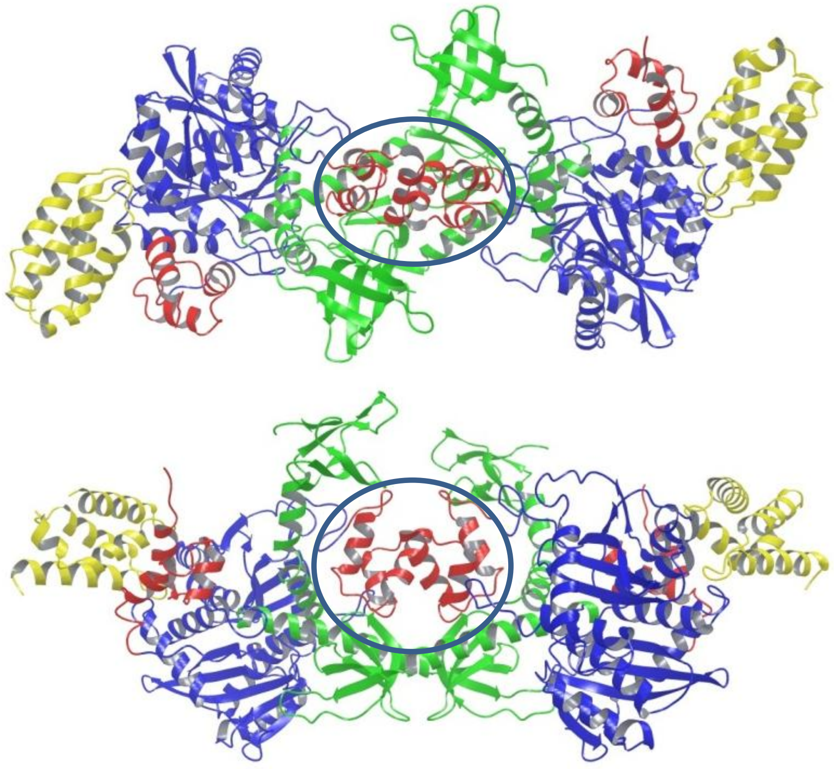


Figure 11: Initial tetramer model, two different views. Clashes between proximal NTDs are encircled. Red: NTDs. Blue: CCDs. Green: CTDs. Two LEDGF IBDs (Yellow) are bound at the high-affinity sites.

First, four monomers were constructed from the different crystallographic subunits using Prime[66] (Methods), the two clashing NTDs (belonging to monomers A and C) were excised from the tetramer and alternative NTD dimer arrangements were generated using the PatchDock[67] and FireDock[67] (bioinfo3d.cs.tau.ac.il/PatchDock) tools for protein-protein docking. Top scoring solutions were visually inspected to identify dimers that could be fitted and connected to the tetramer model. A total of seven dimers with the required geometry were identified, which could be clustered into three conformational sub-families. These were subsequently optimized with Prime side-chain refinement followed by MacroModel[68] minimization (Methods).

The following stage included a complementary approach for removing the N-terminal clashes. The full tetramer model was subjected to a ROSETTA[69] 'Fast Relax' procedure, generating 100 models which were then clustered and visually inspected (Methods). All models generated by ROSETTA were essentially identical in their conformational solution to the N-terminal steric violation. Thus, the dimer of proximal N-terminal domains was extracted from a model representing the largest conformational cluster and added to the pool of conformations generated by docking.

The eight different NTD dimers were then docked into the tetramer model by superimposition onto the original proximal NTDs (Methods), generating eight distinct tetrameric structures (Figure 12). Steric violations were subsequently relieved by removing the IN(186-194) loop from monomers A and C followed by MacroModel minimization and Prime side chain refinement of the newly added domains and their 5Å surroundings (Methods).

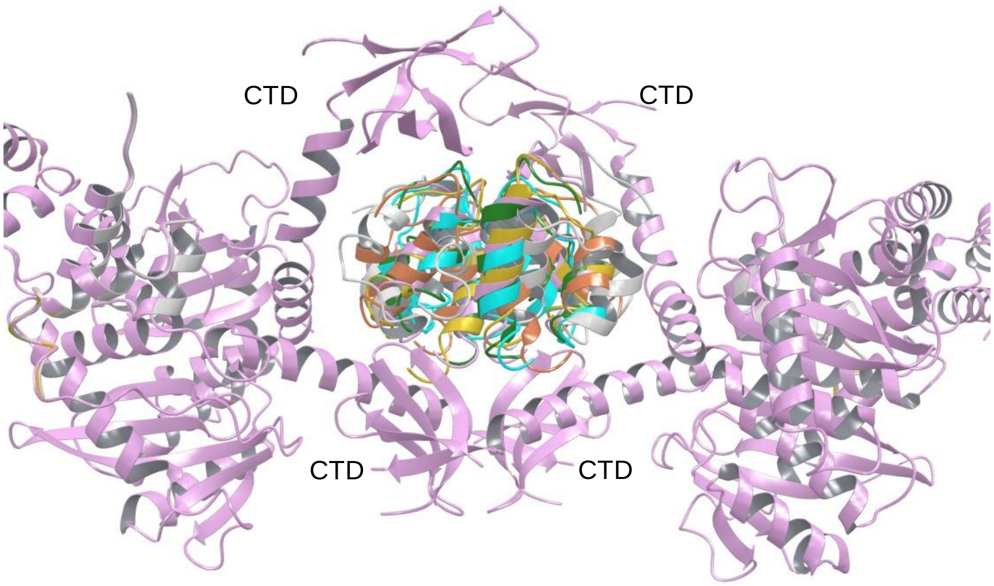


Figure 12: Alternative NTD dimers docked into the initial tetramer model. The model without the proximal NTDs is shown as purple ribbons. The alternative NTD dimer conformations are shown as multicolored ribbons. The tetramer interface is composed of proximal NTDs as well as CTDs.

Each of the eight structures was subsequently used as a template for homology modeling of the full tetramer using MODELLER[70], which allows backbone movements and improved clash removal compared to Prime (Methods). A single initial model was generated for each template structure and subjected to MD refinement of tetramer interface regions in MODELLER (Methods). Of the ten solutions that were generated for each model, a single structure was selected by the Discrete Optimized Protein Energy (DOPE) score[71].

All eight models were subsequently optimized with Prime side chain refinement, which utilizes a more elaborate rotamer library than MODELLER (Methods). Seven of the resulting models had similar Prime energy (~-38,000 kj/mol) while a single model stood out having a lower energy (~-39,000 kj/mol). This model, based on a proximal NTD dimer belonging to the largest conformational cluster of protein-protein docking solutions, was therefore selected as the final model and used for subsequent analysis.

The tetramer in this model is of elongated shape, with two high affinity binding sites for LEDGF at its opposing ends. These are distinguished from the low affinity sites by inclusion of the two distal NTDs (Figure 13), providing an electrostatic anchor for IBD binding[33,36]. Tetramerization is mediated by the two proximal NTDs as well as the four CTDs. However, the CTD contribution is asymmetrical and CTDs from monomers B and D (which also contribute the distal NTDs) are considerably more dominant at the tetramerization interface (Figure 14).

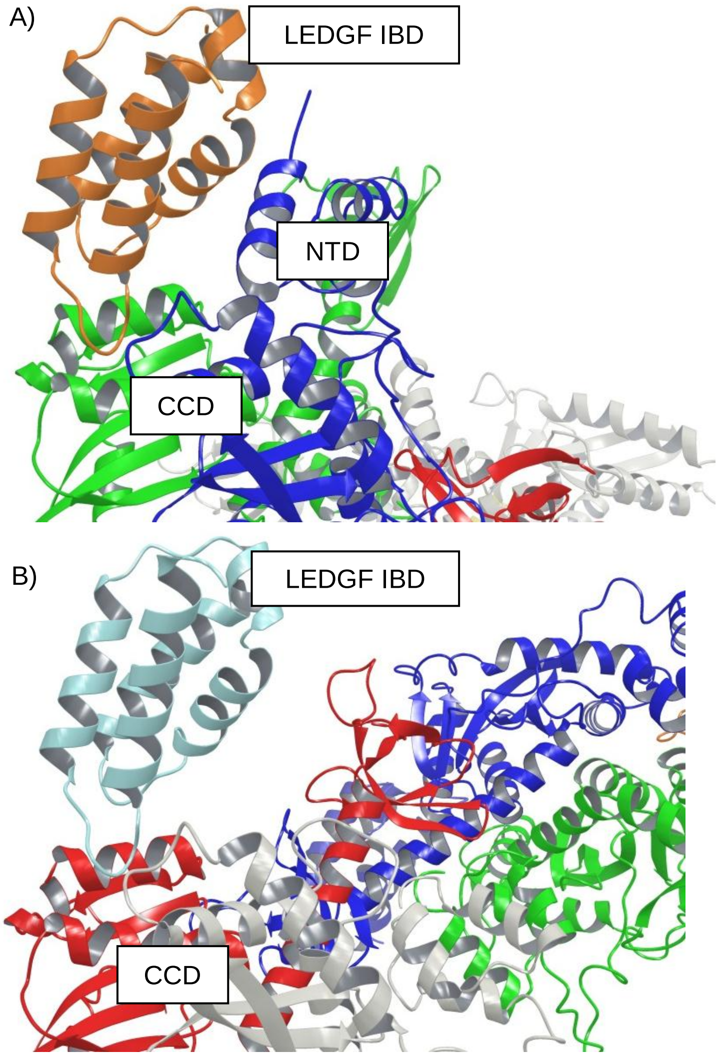


Figure 13: LEDGF binding sites in the final model. A) The high affinity site formed by IN NTD-CCD (blue, green) with bound LEDGF IBD (brown). B) The low affinity site formed by CCD alone (red, grey) with bound IBD (cyan). The interaction between the LEDGF IBD and the IN NTD observed in the high affinity site (A) is clearly absent from the low affinity site (B).

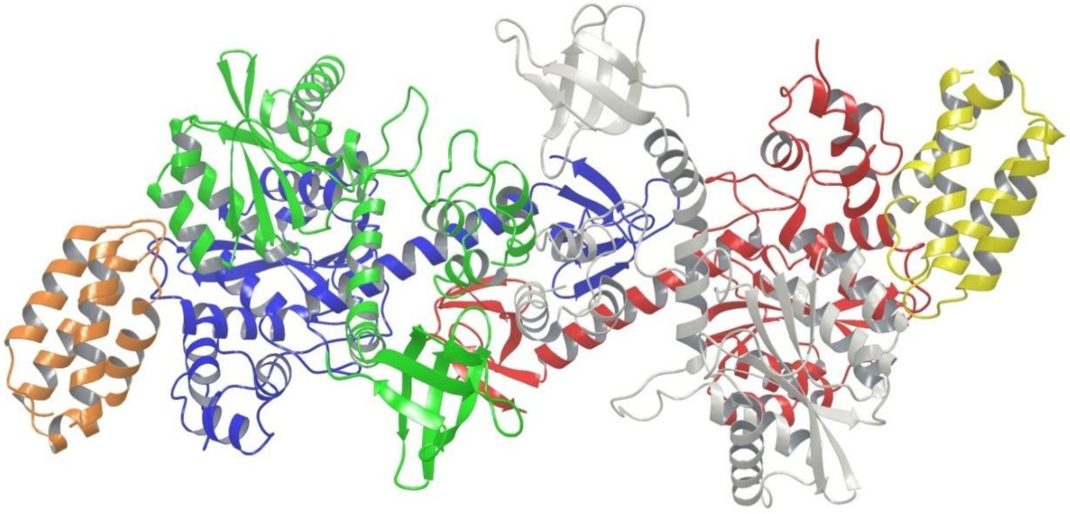


Figure 14: Final tetramer model with LEDGF IBDs (yellow and orange) bound at the high affinity sites. Grey: monomer A, Red: monomer B, Green: monomer C, Blue: monomer D. Monomers B and D contribute the distal NTDs as well as the CTDs that form the main tetramerization anchor. CTDs from monomers A and C largely protrude into the solvent.

4.B.1.B. GNM analysis reveals the allosteric effect of ligand binding

A Gaussian Network Model (GNM[72,73]) calculation was initially performed for three structures: 1) un-bound tetramer; 2) tetramer bound to two LEDGF(361-370) peptides bound at the high affinity sites; 3) tetramer bound to four peptides interacting with the high- and low-affinity sites. These were subjected to a differential analysis as described below in which dynamic properties calculated for the unbound tetramer were subtracted from those calculated for the bound complexes and the effects of the different binding configurations were compared.

The slowest modes of motion extracted by GNM analyses have been shown as relevant to biological function[74,75]. Inspection of the raw GNM results for the three structures revealed that peptide binding does not significantly change the eigenvalue distribution and the distribution of residue fluctuations along the slowest modes of motion is highly similar in all three structures, facilitating comparison between corresponding modes of the different structures.

In order to select the set of slow modes to be used in subsequent analysis, I inspected the distribution of relative mode contributions $1/\lambda_k$, where λ_k is the eigenvalue corresponding to the eigenvector being evaluated (Figure 15). Based on this distribution I applied a cutoff of $1/\lambda_k=1$ and obtained the eight slowest modes, accounting for ~30% of the total motion. First, I looked at the mean square fluctuations of IN residues along each of these modes. Mode 1 relates to fluctuations of the independent monomers with hinges at the interfaces, modes 2-4 relate to motions of the CTDs and the two proximal NTDs and modes 5-8 relate to fluctuations of the CCDs and distal NTDs with clear hinges at the ligand binding sites (Figure 16A). Thus, these modes reflect motions that are relevant for inspection of possible effects of ligand binding on tetramer stability. Mapping the hinges onto the protein structure reveals locations at the ligand binding sites, IN dimerization interface and the tetramerization interface, suggesting that ligand binding might influence the stability of the IN tetramer through modulation of fluctuations around these hinges, which control the cooperative motion defined in the respective modes (Figure 16B). The larger contribution of the distal NTD containing monomers to the tetramerization interface also corresponds to a larger number of interface hinges.

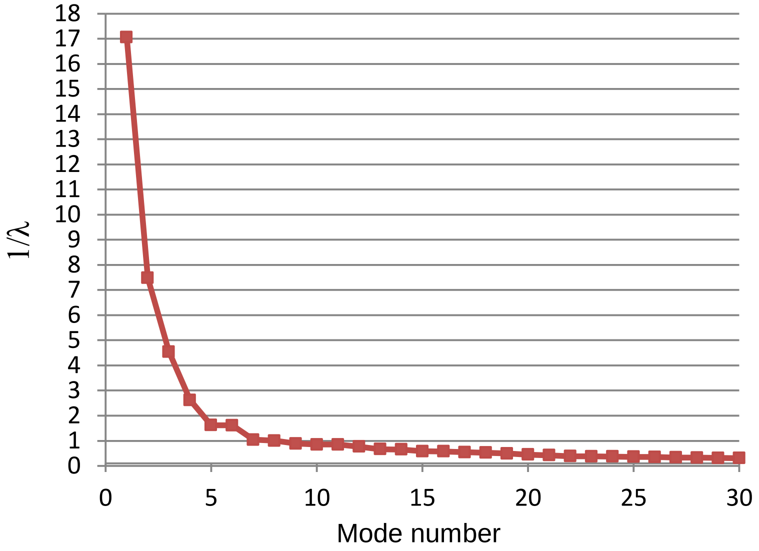
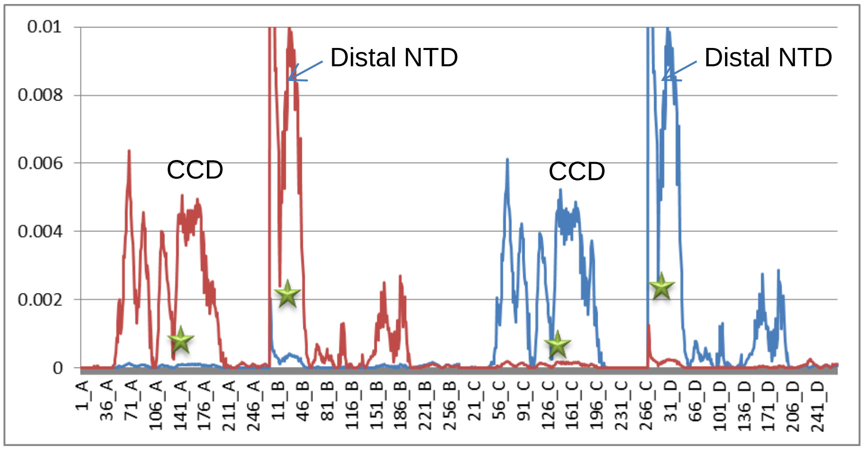


Figure 15. Distribution of 1/ λ values as calculated for the tetramer bound to four peptides. The distribution obtained with only two peptides bound was similar. $1/\lambda_k$ corresponds to the relative contribution of mode k to the total dynamics. I looked closely at all the modes whose $1/\lambda_k \geq 1$, i.e., the first 8 modes.

Modes 1-8 were subsequently averaged and two comparative measures were derived from the GNM properties of the average mode (Methods): 1) differences in amino acid mean square fluctuations as a result of ligand binding; 2) Root mean square deviation of inter-residue cross-correlations between bound and unbound states, measuring the extent of change in the cross correlation of each residue with all other residues as a result of ligand binding. On a side note, hinge regions observed in the individual modes are preserved in the average mode.

A)



B)

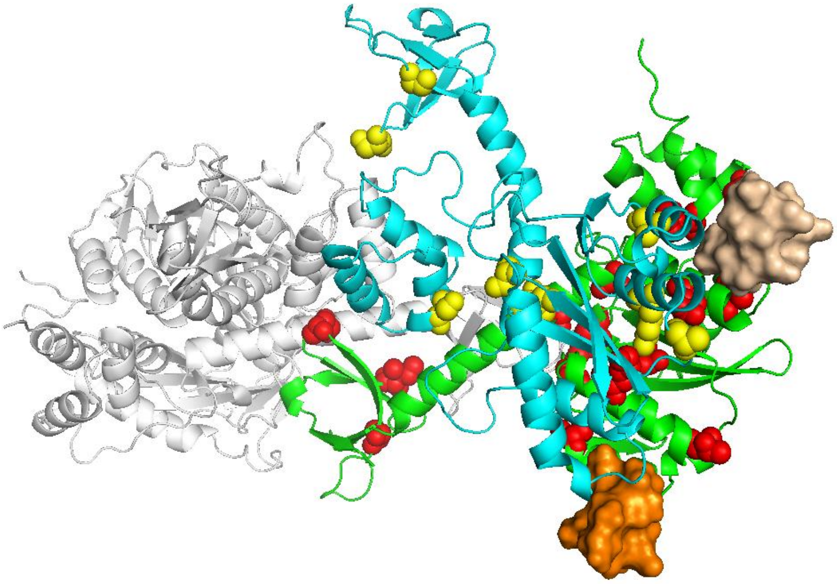


Figure 16: Fluctuations along selected modes and hinge location. A) Fluctuation of tetramer residues along modes 5 (blue) and 6 (red). Green stars mark fluctuation minima (hinge regions) corresponding to LEDGF binding sites, in this case the high affinity sites; B) Hinge residues identified at fluctuation minima of the average mode 1-8. Cyan: monomer A, Green: monomer B, Grey: monomers C and D, Yellow: hinge residues from monomer A (residues 24, 80, 91, 103, 133, 200, 239, 255), red: hinge residues from monomer B (residues 15, 40, 55, 59, 76, 80, 93, 103, 133, 162, 177, 196, 203, 239, 246, 257), orange and wheat surfaces: LEDGF(361-370) bound at the low- and high-affinity sites, respectively. Hinges are located at the ligand binding sites, the CCD dimerization interface and the IN tetramerization interface.

In the differential analysis of mean square fluctuations, peptide binding was found to have a significant effect on tetramer dynamics, increasing fluctuations at the tetramerization interface while reducing the fluctuation of the CCDs (Figure 17). This suggests stabilization through interplay between enthalpic and entropic contributions in which ligand binding sites are rigidified while the entropy at the tetramerization interface increases, thus favoring tetramer formation.

Furthermore, the effect of four bound peptides was significantly larger than that of two peptides bound at the high affinity sites (Figure 17). This suggests that the full inhibitory allosteric effect of LEDGF(361-370) occurs when peptides are bound at all four IBD binding sites. This may indicate that the stoichiometry for IN:LEDGF(361-370) binding (which has not been measured experimentally) is 4:4 rather than the 4:2 ratio measured for IN:LEDGF(IBD).

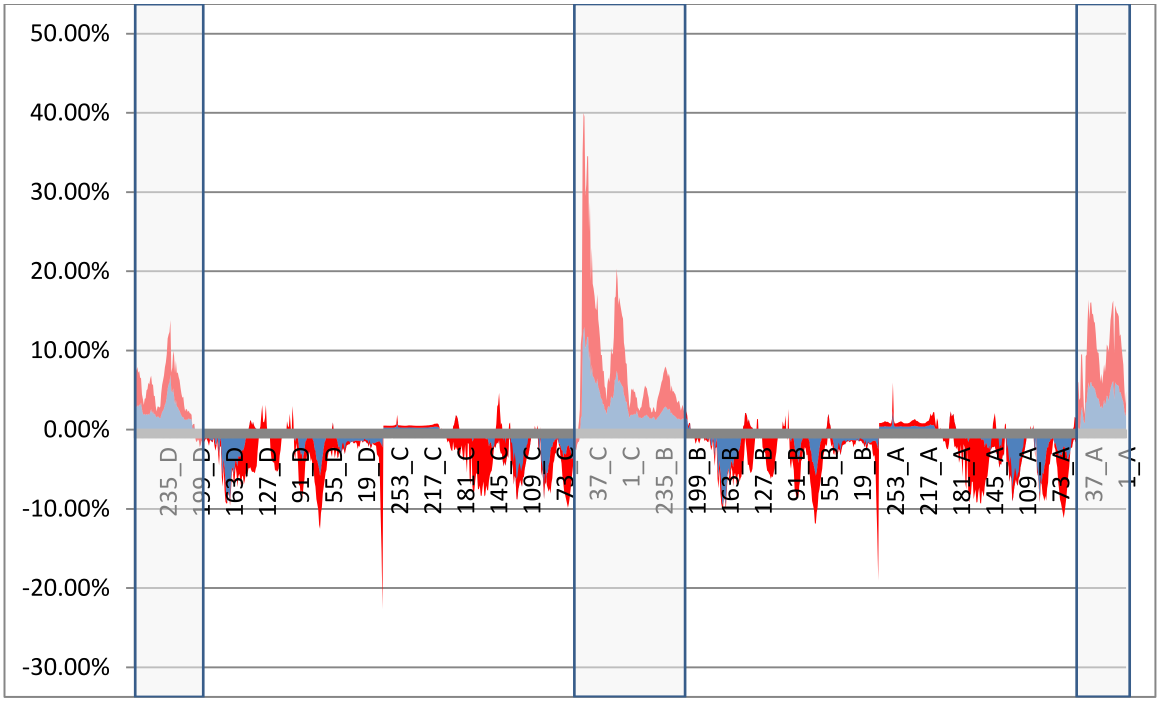


Figure 17: Effect of peptide binding on mean square fluctuations of IN tetramer residues along the average mode 1-8, displayed as percentages of fluctuation of the unbound state. Blue area corresponds to changes following binding of two peptides at the high affinity sites. Red area corresponds to the incremental change following binding of two additional peptides at the low affinity sites. Significant peaks in differences are observed at the tetramerization interface regions (shaded).

In the differential analysis of cross-correlations, an interesting effect was observed. While the effect of two peptides bound at the high affinity sites was mostly limited to the corresponding binding site regions, the effect of four peptides clearly extended into the tetramer interface. To further analyze this effect I performed a similar calculation on a complex with two peptides bound at the low affinity sites. As for the complex with two peptides occupying the high affinity sites, differences in cross correlations were mostly limited to the binding site regions (Figure 18). This implies an unexpected synergistic effect between the two sets of binding sites, suggesting a 4:4 stoichiometry for peptide binding, as does the fluctuations analysis.

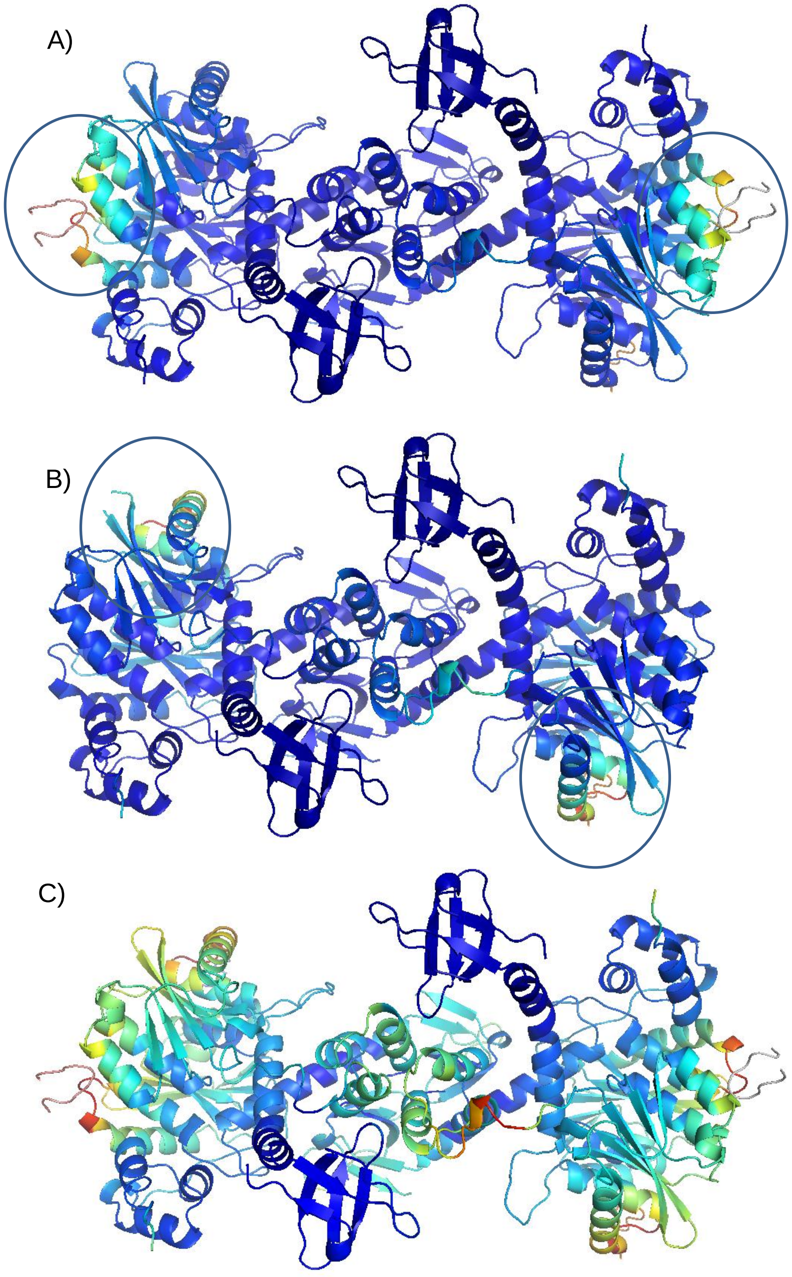


Figure 18: Effects of peptide binding on inter-residue cross-correlations. A) Two peptides bound at the high-affinity sites (circled); B) Two peptides bound at the low affinity sites (circled); C) four peptides bound. Warmer colors represent larger change compared to cross-correlations of the unbound tetramer. The

effect of two bound peptides is mostly limited to the binding site regions. In contrast, the effect of four bound peptides extends towards the tetramer interface, suggesting a synergistic effect between the two pairs of binding sites.

A path of proximal hinges connecting the low affinity and high affinity sites proposes a mechanism by which ligand binding at one site reinforces the conformational effect conducted through the other (Figure 19).

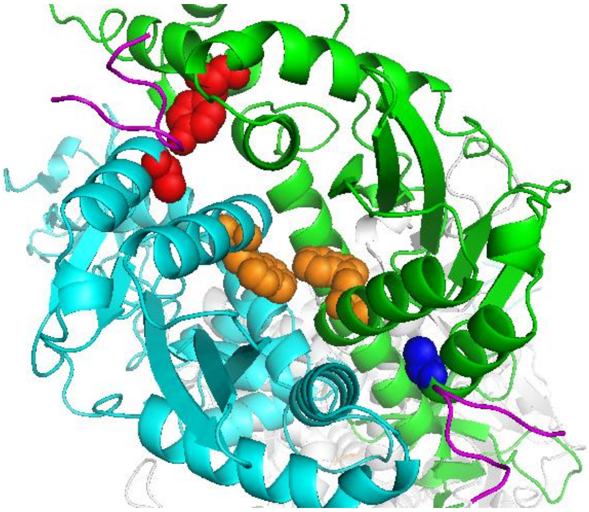


Figure 19: Hinges found in the average mode 1-8 interconnect the high- and low-affinity sites. Green: monomer B, Cyan: monomer A, Red: high affinity site hinges (Y15, A133), Blue: low affinity site hinge (A133), orange: core hinges (W108), magenta: LEDGF(361-370). The core hinges may communicate the binding signal between the low-and high-affinity sites.

The relevance of the present analysis can be challenged by testing two additional peptides with known tetramer stabilization activity. One is an 11-mer Trp-LEDGF(361-370) which is identical in its activity to LEDGF(361-370) (Prof. Assaf Friedler, personal communication) and the other is Trp-LEDGF(365-369), 6-mer with significantly reduced tetramer stabilization activity[31].

Complexes with both peptides were modeled based on the complexes with LEDGF(361-370) and similar calculations were performed which were also based on an average of the slowest modes 1-8 (following inspection of individual modes).

The results for these two additional peptides lend support to the current methodology. The distributions of differences in mean square fluctuations and cross correlation RMSDs obtained for W-LEDGF(361-370) were identical to those of LEDGF(361-370) ($p>0$, two-tailed T-test). In contrast, the effect of the 6-mer peptide was significantly lower than that of the 10- or 11-mers ($p<0.001$, two-tailed T-test), but the effect of four bound 6-mers was still larger than the effect of two bound LEDGF(361-370) peptides, again suggesting that the peptide assumes full potency when bound at all four binding sites.

Finally, I analyzed the binding of LEDGF IBDs. The results suggest that the IBD has a different effect on tetramer dynamics than its short peptide derivatives.

Analysis of IBD binding was based on an average of the ten slowest modes, accounting for ~33% of the motion. These were selected as for the peptide-bound structure, following the application of the $1/\lambda_k=1$ cutoff (Figure 20).

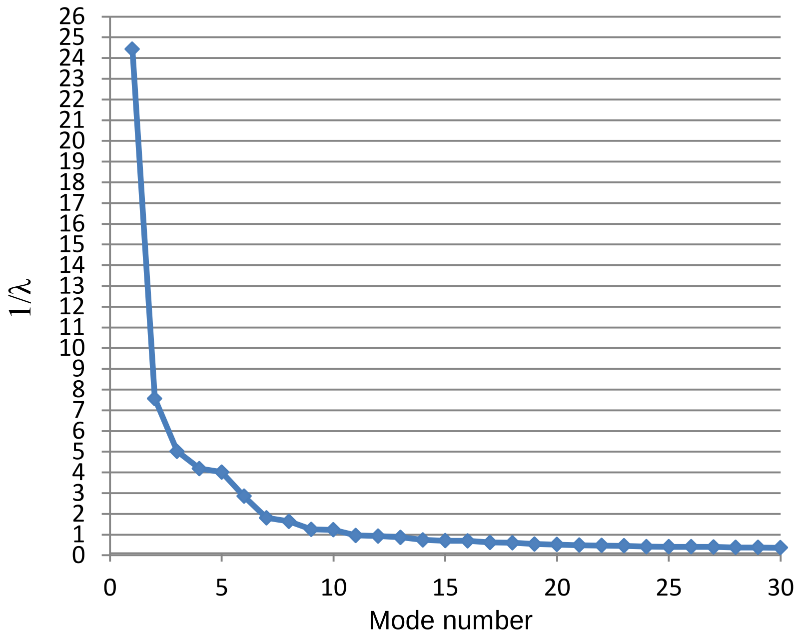


Figure 20. Distribution of 1/ λ values as calculated for the tetramer bound to four LEDGF IBDs. A similar distribution was observed also for the two-bound structures. 1/ λ_k corresponds to the relative contribution of mode k to the total dynamics. A cutoff of 1/ λ_k =1 was applied in selecting the most significant set of slowest modes.

Differential analysis of mean squared fluctuations revealed that the IBD binding at the high affinity sites significantly reduces the fluctuation of the distal NTDs that constitute an important part of the LEDGF high affinity binding sites (Figure 21). Interestingly, the incremental effect of ligands bound at the low affinity sites is less prominent for the IBDs than for IBD-derived peptides. In contrast to the modulation of IN fluctuation by peptide binding, IBD binding mostly reduces amino acid fluctuation, suggesting a largely enthaplic effect on tetramer assembly, perhaps through stabilization of the dimer conformation required for tetramerization.

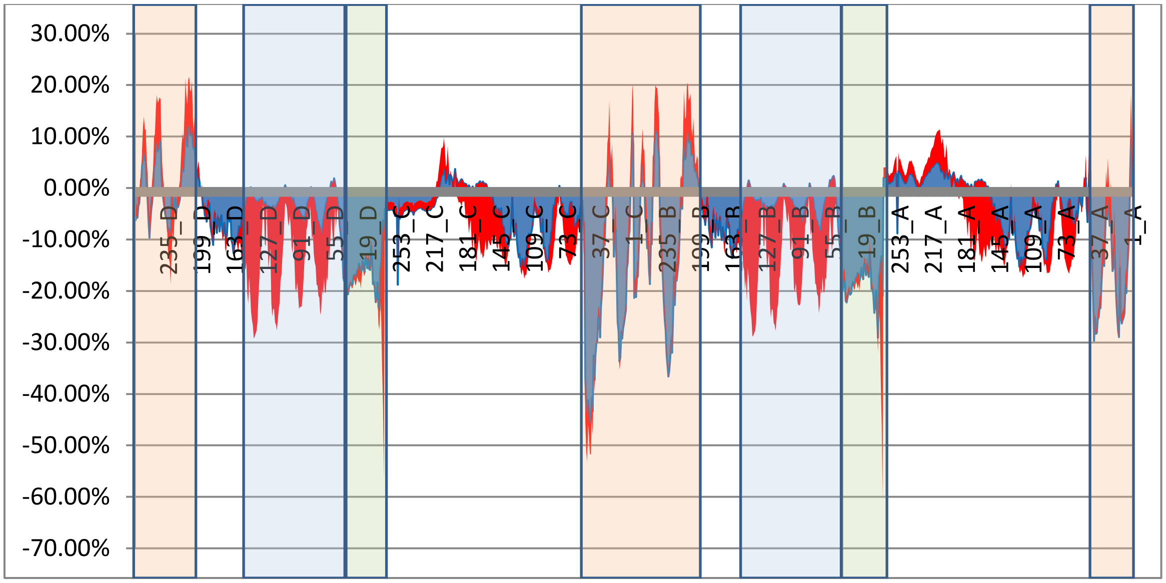


Figure 21: Effect of IBD binding on mean square fluctuations of IN tetramer residues along the average mode 1-10, displayed as percentages of fluctuation of the unbound state. Blue area corresponds to changes following binding of two IBDs at the high affinity sites. Red area corresponds to the incremental change following binding of two additional IBDs at the low affinity sites. Main peaks in fluctuation differences are observed in the CCD regions surrounding the low affinity sites (blue shaded) rather than tetramerization regions (brown shaded) as observed for peptide binding. IBD binding at the high affinity sites significantly decreases fluctuations is the distal NTDs (green shaded).

Differential analysis of cross-correlations also suggests that two-bound IBDs have a similar effect on the tetramer interface to that of four bound IBDs (Figures 22, 23), in contrast to the synergistic effect observed for peptide binding.

This analysis of IBD binding suggests that two IBDs bound at the high affinity sites may be sufficient for tetramer stabilization, in agreement with the experimentally determined stoichiometry and SAXS results[33].

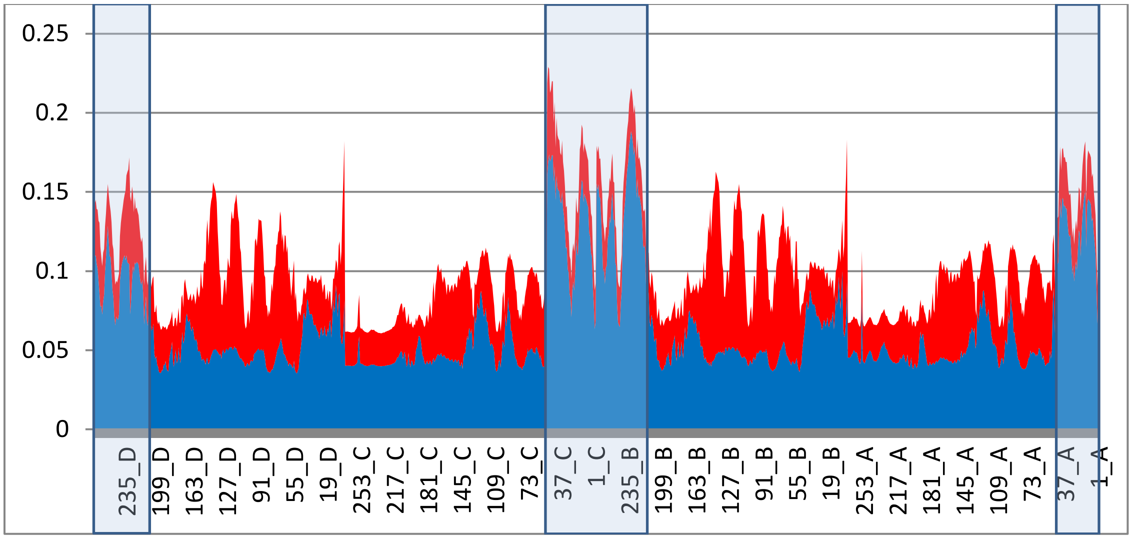


Figure 22: Effect of IBD binding on inter-residue cross-correlation. Blue: effect of two IBDs bound at the high affinity sites. Red: Incremental effect following binding of two additional IBDs at the low affinity sites. In contrast to peptide binding, the smallest incremental effect is observed at the tetramerization interface (shaded).

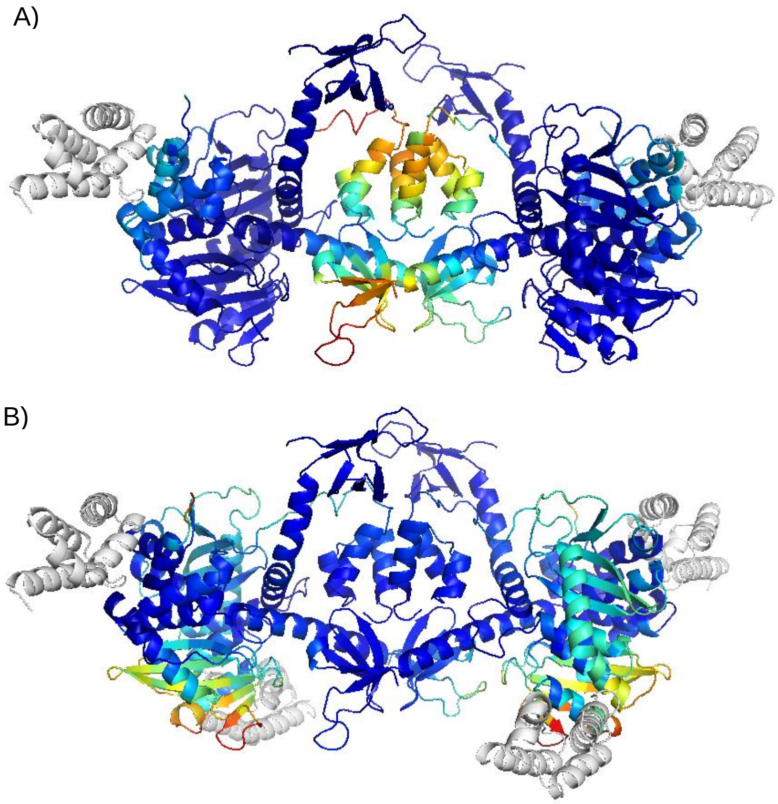


Figure 23: Effects of LEDGF IBD binding on inter-residue cross-correlations. A) Effect of two IBDs bound at the high-affinity sites compared to unbound state; B) Incremental effect of IBD binding at the low affinity sites. . Grey: Bound IBDs. Warmer colors represent larger changes. Binding at the low affinity sites has a mostly localized effect, suggesting that two LEDGFs bound at the high affinity sites could be sufficient to stabilize the tetramer.

4.B.2. Discussion

I have used GNM analysis to investigate the stabilizing effect of LEDGF and peptides derived from LEDGF on the inactive IN tetramer using two measures calculated from the slowest modes of motion: 1) differences in amino acid mean square fluctuations upon binding; 2) RMS deviations in amino-acid cross-correlations upon binding.

The results showed that the effect of LEDGF(361-370) binding on tetramer dynamics extends beyond the binding site region into the tetramerization interfaces. Furthermore, the results suggested that occupation of all four binding sites may be required for the peptides to gain full potency, with a synergistic effect between binding at the high- and low-affinity sites.

In contrast, the results suggest that binding of two LEDGF IBDs at the low-affinity sites does not significantly augment the effect of IBD binding at the high-affinity sites, in agreement with the experimentally observed 4:2 (IN:IBD) stoichiometry. This may relate to the more dominant role of monomers B and D in forming the dimer of dimers and the role of the distal NTDs in transmitting the allosteric signal to the protein-protein contact interfaces, as reflected by the number of interface hinges contributed by monomers B and D. While the IBD forms multiple interactions with the distal NTDs, the bound peptides are not in direct contact with these domains and may require the additional binding event to accumulate the required potency.

The GNM results also suggest a different mechanism for tetramer stabilization by LEDGF IBD and its derived peptides. Peptide binding is observed to reduce fluctuations of the CCDs while increasing fluctuation of the tetramerization interface. In contrast, IBD binding, which involves a greater contact surface including the distal NTDs, appears to mostly reduce fluctuations. This suggests entropy driven stabilization by the peptides compared with an enthalpy-driven stabilization by IBD binding.

Results for two additional peptides lend support to the physical relevance of the measures used for assessing complex stabilization by ligand binding. As expected from experimental results, an identical effect was observed for LEDGF(361-370) and Trp-LEDGF(361-370). In contrast, the shorter Trp-LEDGF(365-369) was found to have a significantly smaller effect according to both measures, in agreement with its reduced ability to stabilize the tetramer. The 6-mer is of lower affinity than the 10-mer[31] and lacks two residues important for 10-mer affinity, S362 and V370. Both residues do not interact directly with IN. S362 forms an intra-molecular hydrogen bond, potentially contributing to the stability of the peptide conformation required for binding, and V370 stacks on top of L368, which is in contact with binding site residues T124 and A128. While the latter two are not hinge residues themselves, they are on the same short helical stretch as residues 131-133, which appear as hinge regions in the low frequency modes. V370 may thus be important for reinforcing the allosteric effect of the bound peptide.

This latter result bears additional significance. A recent publication described small molecule inhibitors of the IN-LEDGF binding[76]. In addition to inhibition of IN by LEDGF displacement, these molecules have also been found to moderately inhibit IN enzymatic activity, probably through an allosteric mechanism.

These compounds overlap well with the di-peptide LEDGF(365-366) (Figure 24), suggesting an allosteric mechanism similar to that of the LEDGF-derived peptides.

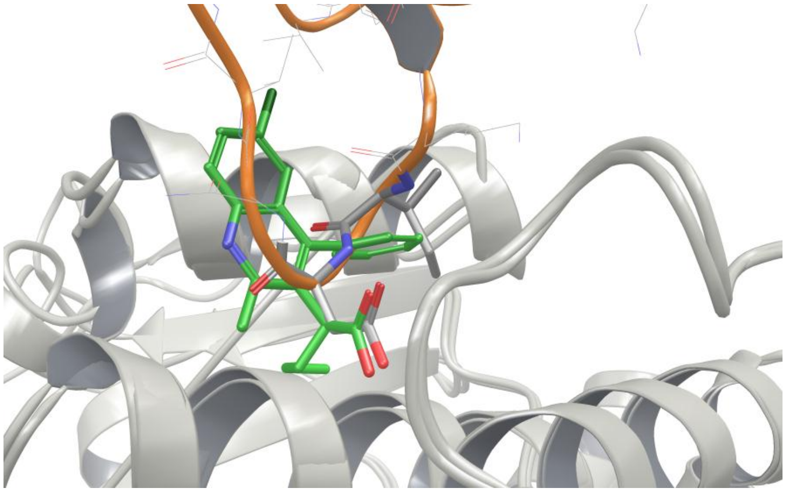


Figure 24: Overlap between small molecule inhibitor and LEDGF(365-366). The inhibitor bound structure 3LPU[76] was superimposed onto the LEDGF bound structure 2B4J[40] using Maestro[77]. The IN CCD and LEDGF structures are represented by grey and orange ribbons, respectively; the bound inhibitor is shown in sticks representation with green carbons; the IBD core residues I365 and D366 are depicted in sticks representation with grey carbons.

This observation reflects on the second part of my IN-LEDGF project, in which I performed structure-based screening for small molecule inhibitors of the IN-LEDGF interaction. Compounds purchased for *in-vitro* testing will be evaluated in an integration assay without LEDGF to measure their allosteric affect as well as the same assay with LEDGF present to assess their ability to inhibit LEDGF potentiation of IN activity (Assays performed in the lab of Prof. Assaf Friedler, the Hebrew University).

Being highly rigid, the reported small molecule inhibitors as well as the ones that I have designed, do not suffer from the limited conformational stability of the 6-mer peptide and thus may not require the additional stabilization observed in the larger 10-mer. In this case, high affinity may be sufficient for obtaining a potent allosteric effect. Still, increasing compound size to improve its overlap with LEDGF(361-370) may be a viable approach for increasing its allosteric effect.

Finally, the differential GNM analysis introduced in this work may be used to investigate effects of ligand binding in additional systems.

4.C. Project 3:

4.C.1. Results

4.C.1.A. Binding Site Analysis

The recent crystal structure 2B4J depicts the interaction between an IN dimer and two LEDGF units[40]. However, this conformation of the LEDGF binding site may not be ideally adapted for interaction with small molecule inhibitors. Thus, I started by analyzing the LEDGF binding region in a variety of IN crystal structures, including both holo- and apo-structures. This analysis encompassed a total of 34 binding sites in 17 crystal structures of resolution under 2.5Å. Following protein preparation (Methods), binding site druggability was evaluated using the SiteMap program[78]. Interestingly, several apo structures (especially 1BHL[79]) as well as the 1HYV co-crystal structure with Tetraphenyl-Arsonium (TTA[80]) obtained DScore[81] values superior to the 2B4J structure (DScore>1, interpreted druggable, as opposed to DScore<=0.8, interpreted as undruggable), suggesting that these binding site conformations are significantly more suitable for small-molecule drug-design.

Computer-aided drug-design of protein-protein interaction inhibitors may benefit from consideration of binding site flexibility. Therefore, I assessed binding site flexibility and the potential need to incorporate protein flexibility into the drug discovery process by superimposing the various binding site representations and visual analysis of the SiteMap results.

This analysis revealed that flexibility in the LEDGF interaction site is mostly limited to side chain motions of non-conserved solvent exposed residues at the protein surface (Q95, E170, Q168). In contrast, the core of the binding site is more rigid, and as revealed by SiteMap interaction maps, largely involves interactions with backbone atoms (Figure 25). This suggested that in this particular case incorporating binding site flexibility into the screening process would not provide a significant advantage over a rigid receptor representation.

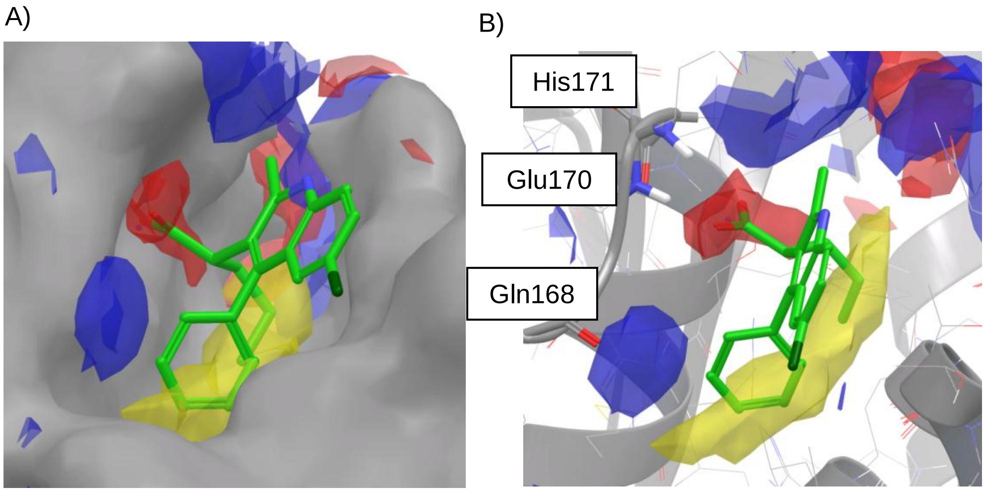
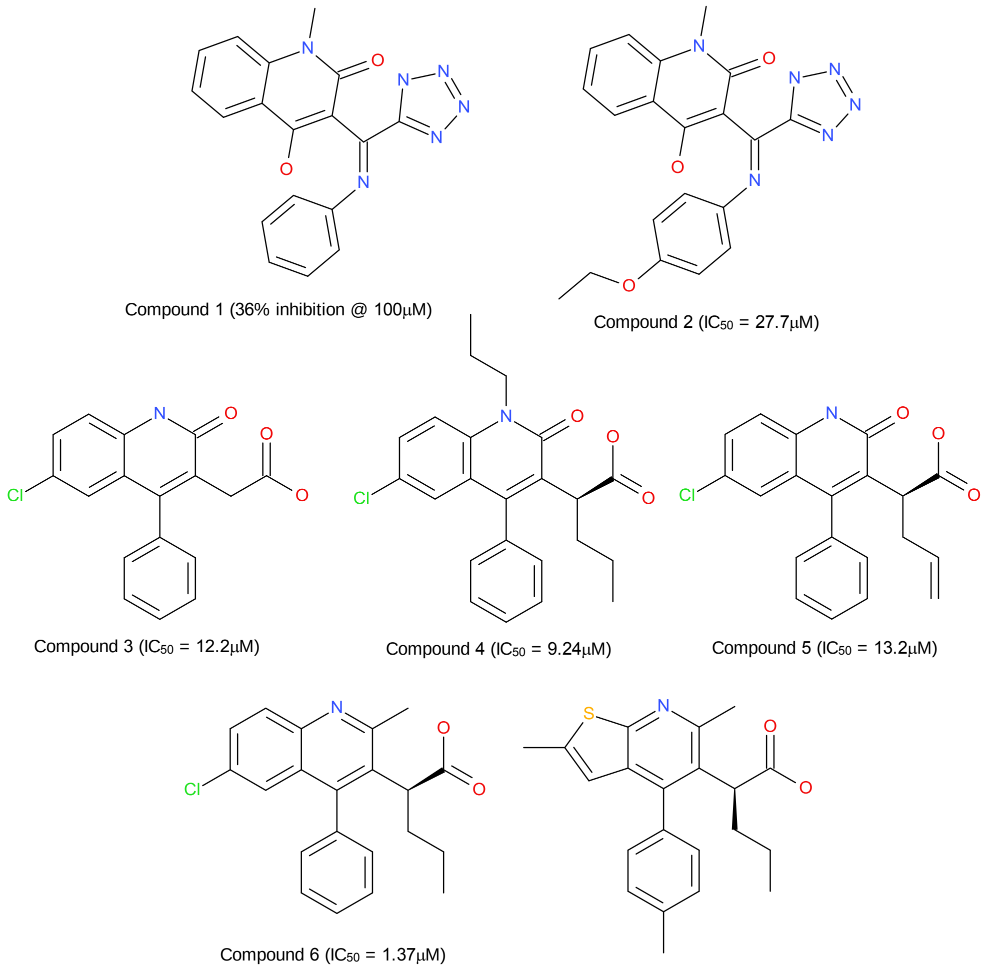


Figure 25: **SiteMap analysis of the LEDGF binding site in the 3LPU co-crystal structure**. Yellow: hydrophobic region, Red: Hydrogen bond acceptor region, Blue: Hydrogen bond donor region. The molecular surface of the IN dimer is depicted in Grey. A) The core region of the binding site is composed of an elongated hydrophobic region accommodating the alkyl side chain and phenyl ring of the co-crystallized inhibitor, a hydrogen bond acceptor region occupied by the carboxylic acid and a hydrogen bond donor region which is apparently unoccupied. The rest of the hydrophilic interaction regions originate from solvent exposed side-chains of non-conserved residues. B) The main contributions to the polar interaction regions are from the backbone atoms of IN residues His171, Glu170 and Gln168. The side of the non-conserved Thr174 also contributes to the hydrogen bond acceptor region.

At this point a paper was published by Debyser and co-workers[76], revealing co-crystal structures of IN with small molecule inhibitors of LEDGF binding. The paper describes seven related compounds with IC_{50} values ranging from 0.58μM to 36% inhibition at 100μM (Figure 26). The least active compound was obtained by virtual-screening of 200,000 compounds, utilizing the 1HYV co-crystal structure with TTA. Two improved inhibitors (IC_{50} values of 27.7μM and 12.2μM) were subsequently found by a similarity search, and 4 additional analogs of the most active compound (IC_{50} under 15μM) were finally obtained by medicinal chemistry.



Compound 1 (36% inhibition @ 100µM)

Compound 2 (IC_{50} = 27.7µM)

Compound 3 (IC_{50} = 12.2µM)

Compound 4 (IC_{50} = 9.24µM)

Compound 5 (IC_{50} = 13.2µM)

Compound 6 (IC_{50} = 1.37µM)

Compound 7 (IC_{50} = 0.58µM)

Figure 26: **Reference compounds** recently described by Christ et al.[76].

This work provided additional co-crystal structures for evaluation, which are already adapted for small molecule binding, as well as several active compounds with which to calibrate the virtual screening process. Both crystal structures, 3LPT (co-crystal with the 12.2µM compound 3) and 3LPU (co-crystal with the 1.37µM compound 6), were added to the binding site analysis obtaining DScores comparable with the top binding sites so far (data not shown). Since these structures are already adapted to binding drug-like compounds, both should provide improved templates over the previously analyzed structures. Among the two, the 3LPU complex with the more active analog scored better in SiteMap and was thus selected as a template for virtual screening.

4.C.1.B. Development of the screening protocol

Compounds from Christ et al. were initially docked into 3LPU using Glide[82,83] (Methods) in order to assess docking performance. Both weakly active analogs 1 and 2 (Figure 2) were excluded from the calibration process. Being structurally distinct from the carboxylic-acid derivatives co-crystallized in 3LPU and 3LPT, these compounds docked differently and the predicted binding modes could not be validated using the currently available experimental information.

The first observation following this initial experiment was that docking results improve considerably when applying constrains based on binding site interactions. As clearly seen in Figure 1, the main polar interaction points in the binding site core are the backbone amine groups of Glu170 and His171 and the Gln168 backbone carbonyl. In the 2B4J co-crystal structure with LEDGF, the Gln168 carbonyl forms a hydrogen bond with the backbone of LEDGF residue Ile365 while the two donor groups bind the carboxylic acid of Asp366[40]. The latter are also important for the recognition of carboxylic acid inhibitors, as observed in 3LPU and 3LPT. Requiring the fulfillment of at least one of the Glu170/His171 hydrogen bonds improved docking results considerably over unconstrained docking, as assessed by comparison of the docked poses to the bound conformations observed in 3LPU and 3LPT. This may be due to the shallow binding site geometry and the small number of specific interactions formed by the inhibitors (only two hydrogen bonds), constituting a relatively “weak signal” for docking. Still, even after the addition of this constraint, Glide was unable to consistently produce the correct binding mode as the top ranked solution for all compounds. In addition, evaluation of different compound preparation protocols revealed significant sensitivity to initial conformation with detrimental results if the “wrong” preparation protocol was selected (data not shown). Based on discussion with Schrödinger support scientists, it seems that the latter could be attributed to the absence of a strong “docking signal” as well as conformational coverage limitations in Glide.

My solution to both problems was to devise a protocol in which docking is preceded by a structure-based pharmacophore search and followed by a structure-based pharmacophore filter, as described below.

A crude structure-based pharmacophore hypothesis was first constructed in the 3LPU binding site using Phase[84], based on the bound conformation of compound 6. This has been designed to include the chemical flexibility required for identification of novel inhibitor chemotypes in screening. Two hydrophobic-aromatic (HR) features were mapped onto compound 6 phenyl ring (HR1) and propyl side-chain (HR2), coinciding well with the hydrophobic region revealed by SiteMap. HR1 can clearly occupy both aromatic and hydrophobic groups as exemplified in the compounds 3-7 phenyl ring and LEDGF Ile365 side chain. By symmetry, there is no reason to prevent aromatics from matching HR2, despite the current lack of an experimental example. Two hydrogen bond acceptor (A1, A2) features were placed on the carboxylic acid oxygen atoms and a combined acceptor-negative (AN) feature was placed at the geometric center of the carboxylic acid group (Figure 27, Methods). While HR1 was defined as a necessary "must match" feature, HR2 was set as optional in order to match compound 3 which is active but does not contain the respective alkyl side-chain. As there is no apparent structural requirement for a negatively charged group in place of the carboxylic-acid, partial matches were allowed to the (A1, A2, AN) group and planar hydrogen bond directionality was not imposed, allowing a wide range of hydrogen bond acceptor atom types to match. Receptor-based excluded volumes were then added to the pharmacophore hypothesis to further focus the matching process to solutions that sterically fit the binding site.

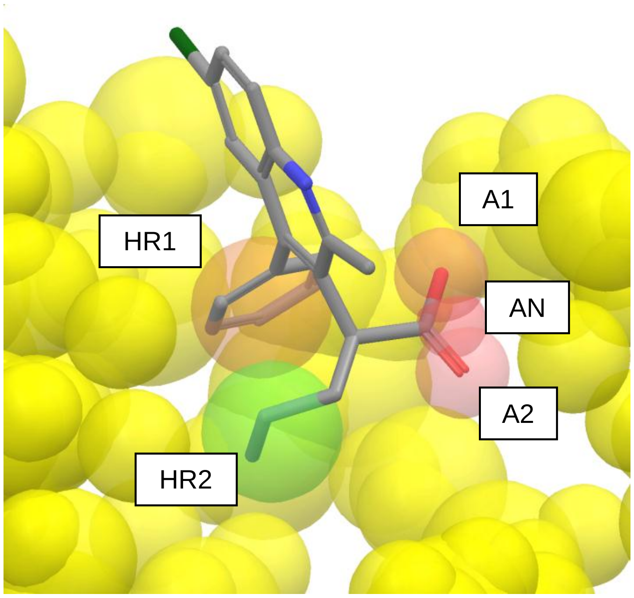


Figure 27: **Structure-based pharmacophore**. Pharmacophore features are depicted as spheres centered on the corresponding chemical groups of the co-crystallized compound 6. HR: Hydrophobic or Aromatic, A: Hydrogen bond Acceptor, AN: Acceptor or Negatively charged. Sphere radii correspond to the respective feature matching tolerance values and are in scale with the inter-atomic distances in the molecule. Receptor-based excluded volumes are depicted as yellow spheres.

Matching compounds 3-7 to the structure-based pharmacophore (pharmacophore screening) resulted in binding modes that agree with the co-crystallized conformations of compounds 3 and 6. These poses were subsequently used as input for Glide docking, which could be viewed as biasing the Glide conformational search towards solutions matching the binding site, thus potentially avoiding the observed sensitivity to initial conformation. The same structure-based pharmacophore was reapplied as a filter to the top five docking solutions of each compound, selecting a single binding pose which best agrees with the interaction determinants derived from SiteMap analysis and the Christ et al. structure-activity relationship (SAR), thus potentially avoiding the "weak signal" effect.

This process, which does not rely on chemical similarity criteria and is generic enough to be used for screening, resulted in binding modes for compounds 3-7 that are in good agreement with experiment (assuming that compounds 4,5 and 7 adopt similar binding modes to 3 and 6, which is a reasonable assumption since they are all close-in analogs, Figure 28).

In addition, approximate ΔG scores calculated with Prime MM-GBSA[66] correlated reasonably well with experimental activities (R^2=0.63), lending support to credibility of the selected binding modes and the ability of scoring functions to assess experimental affinity, at least for this series of compounds.

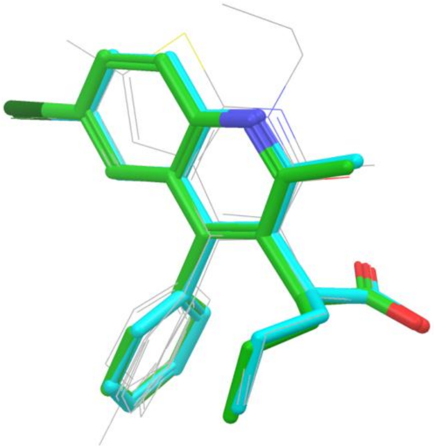


Figure 28: Top ranked binding mode for known inhibitors docked using the combined screening protocol, superimposed onto the co-crystallized compound 6. Near perfect overlap is obtained between the co-crystallized compound 6 (green carbons) and the top scoring binding mode of the docked compound 6 (blue carbons) and the related analogues all attain similar binding modes (grey carbons, thin lines).

4.C.1.C. Screening

A screening library focused on chemical space relevant for drug like protein-protein interaction inhibitors was prepared as described in Methods and subjected to the screening protocol described above. Subsequently, a Glide Docking Score cutoff of -5.8 was applied as a lower bound, based on the score range obtained for the known inhibitors used for process calibration ([-5.98, -6.57]). 1,077 compounds survived this initial filtration, corresponding to an enrichment factor of 789. This measure cannot be used as a reliable predictor of screening success due to the limited diversity of the known compounds.

To enable further library filtration, the remaining compounds were re-scored with the following scoring functions: ligScore1 Drieding, ligScore2 Drieding, PLP1, PLP2, JAIN, PMF, PMF04[85] and Prime-MM-GBSA ΔG approximation score (Methods). Of the two inactive and weakly active compounds described in Christ et al, only compound 1 survived the screening process. Therefore, this compound was used as a negative control to evaluate scoring function performance. Subsequently, PMF, PMF04 and PLP1 which ranked compound 1 better than at least one of the active compounds 3-7, were excluded from the analysis.

For the remaining scoring functions, score cutoffs were determined as for the Glide Docking score and a rank-by-vote consensus score was calculated.

All compounds with Consensus Score > 0 were clustered based on molecular similarity using Canvas[86] as described in Methods. Cluster representatives were visually inspected in the binding site, and considered for *in-vitro* testing based on the following criteria: 1) Favorable receptor-ligand interactions versus unfavorable interactions assessed by comparison to the SiteMap interaction map and evaluation of hydrogen bond geometry. At least one hydrogen bond was required with the Glu170 and His171 backbone NH groups, as described above and compounds with an additional hydrogen bond with the Gln168 backbone CO were favored. Proper filling of the hydrophobic region at the binding site core was required and a minimal bound was set according to compound 3 which does not match the HA2 pharmacophore feature; 2) Conformational strain estimated by conformational analysis of the free ligand (Methods) and comparison of the bound conformation to the solution ensemble. Compounds with highly strained conformations were discarded; 4) Compound flexibility: flexible compounds were deprioritized to avoid large entropic penalties upon binding; 5) Diversity: diversity compared to the known inhibitors was required as well diversity with the selected set of compounds in order to cover as much of the relevant chemical space as possible within a small set of selected representative compounds; 6) Chemical liabilities were considered and compounds containing groups that are known to be chemically unstable were discarded.

The final set of selected compounds contained 86 diverse compounds that could be roughly divided into two classes: carboxylic acid derivatives and non carboxylic-acid derivatives.

4.C.1.D. Selected compounds

Carboxylic acid derivatives: 40 of the selected compounds were carboxylic acid derivatives. Reassuringly, the known inhibitor compound 3 (Asinex BAS 02073592) was blindly identified as one of the screened compounds. All the selected compounds are chemically novel with respect to the know inhibitors. Representative structures and predicted binding modes of selected compounds are shown in Figure 29.

Non carboxylic-acids: Of these, one was charged, placing a tetrazole in place of the carboxylic acid and the rest had between one and two hydrogen bonds with the Glu170 and His171 backbone NH groups and two compounds had an additional hydrogen bond with the Gln168 CO. Representative structures and predicted binding modes of selected compounds are shown in Figure 30.

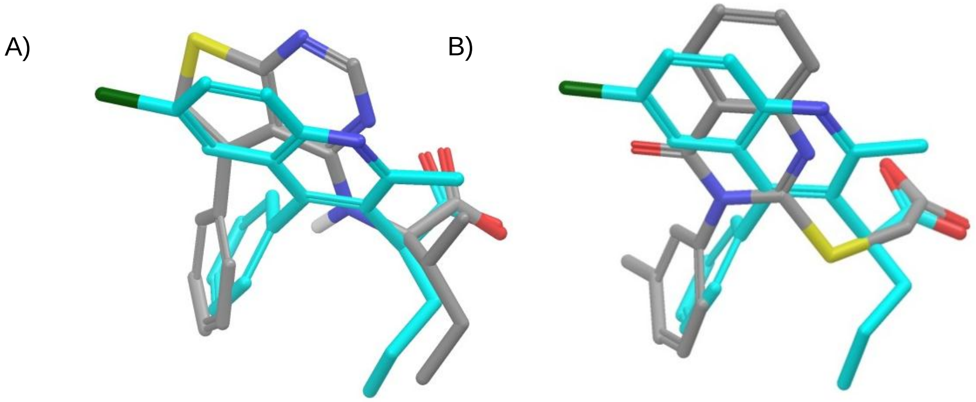


Figure 29: Docked conformation of two carboxylic acids selected for testing (grey carbons) superimposed onto the co-crystallized compound 6 (Cyan carbons). A) Asinex BAS 4913249; B) Asinex ASN 4755011.

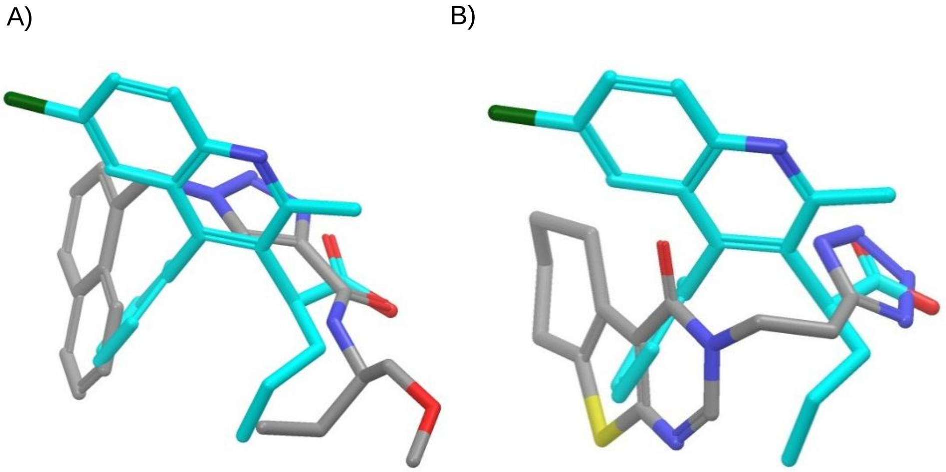


Figure 30: Docked conformation of two non-carboxylic acid compounds selected for testing (grey carbons) superimposed onto the co-crystallized compound 6 (Cyancarbons). A) Chembridge 48371098; B) Chembridge 41865676.

4.C.1.E. *In-Vitro* results

Out of the 86 selected compounds, 45 were tested so far. These include 40 carboxylic acids, one tetrazole and four uncharged compounds. Table 4 shows single concentration data obtained at 1:1000 enzyme-to-compound ratio for 31 out of the 45 compounds. The

14 remaining compounds were excluded due to high standard deviations. The reference compounds used in this assay are compounds 1 and 2 from the Christ set discussed above and a close-in analog of compound 3 from this set. While compound 3 is a 2-(6-chloro-2-oxo-4-phenyl-1H-quinolin-3-yl)acetic acid with IC_{50}~12μM[76], the analog, 2-[6-methyl-4-(1-methyleneallyl)-2-oxo-1H-quinolin-3-yl]acetic acid, bears a solvent-facing chloro substitution which most probably does not affect compound potency. These were the only commercially available compounds that could be used as references at the time.

A small number of compounds have also been tested in the presence of LEDGF (1:1 enzyme to LEDGF molar ratio), showing similar or increased inhibition (Figure 31). This suggests a different balance between direct inhibition and LEDGF displacement.

Interestingly, one compound, BAS 00917565, potentiated IN activity in the direct inhibition assay.

4.C.1.E.1. Estimating potential active-site inhibition

The original hypothesis was that compounds binding at the LEDGF binding site will inhibit IN function by displacing LEDGF and possibly also allosterically, by stabilizing an inactive conformation of the IN tetramer (see Project 2 Discussion). However, there is another option that must be addressed given the enzymatic inhibition obtained *in-vitro*, which is that the compounds also function as active site inhibitors of IN.

To test this hypothesis, the 86 selected compounds as well as the Christ set were matched to a structure-based pharmacophore (SBP) based on the co-crystallized structure of Raltegravir bound to prototype foamy virus (PFV) integrase. The known cross-reactivity between HIV-1 and PFV encompassing at least two different chemotypes, Raltegravir and Elvitegravir[87], suggests that the PFV-bound conformation of Raltegravir may be similar to its HIV-1 bound conformation. This is necessarily true for the rigid planar metal chelating core structure, which clearly demonstrates the co-planarity required for the metal-chelating atoms. Therefore, the SBP was constructed to include the core metal-coordinating pharmacophore[39] and co-planarity[39,88] as well as volume overlap with Raltegravir were used in evaluating matching compounds.

Several compounds from the Christ set (compound 3, 4 and 5 and to a slightly lesser extent compounds 1 and 2) as well as nine of the 27 screening compounds that showed non zero inhibition activity were found to match the Raltegravir-based SBP (Table4, Figure 32). This includes two of the nine compounds with inhibition greater than 80%. The rest were either lacking one of the lone pair donors required for metal chelation or incapable of placing them at the required coplanar orientation.

Table 4. Percent IN inhibition of 31 compounds out of the 45 that were tested

Vendor	Catalog Number	Inhibition of IN's catalytic activity (%)	Normalized STDV	Potential active site inhibitor*
ChemBridge	5375194	100	3	
ChemBridge	5152889	100	1	
ChemBridge	5524250	100	8	
ChemBridge	5403598	98	2	
Asinex	BAS 01507869	94	5	Y
ChemBridge	5103434	91	2	Y
Asinex	BAS 01814325	83	6	
ChemBridge	61120235	83	7	
ChemBridge	5211827	80	2	
ChemBridge	41865676	75	1	
Asinex	BAS 00537204	74	16	
ChemBridge	5190214	73	8	Y
ChemBridge	5401767	68	6	
Asinex	BAS 02179070	67	3	Y
ChemBridge	5160149	67	10	
Asinex	BAS 07330949	62	3	Y
Asinex	BAS 11150812	62	7	Y
Specs	AF-399/15284149 (**)	60	16	
Asinex	BAS 00282908	57	18	Y
ChemBridge	5211829	57	13	
ChemBridge	5789062	52	4	
Asinex	BAS00347768	46	6	
ChemBridge	5213900	45	12	
Asinex	BAS 04913249	37	11	
Asinex	BAS 03754363	22	9	Y
Asinex	ASN 04755011	19	11	
Specs	AE-406/41056636 (***)	17	7	
Asinex	BAS 00692370	16	2	
Asinex	BAS 09659192	11	2	
Asinex	BAS 01556676	-6	2	Y
Asinex	BAS 00917565	-74	18	

* Estimated by fitting to a Raltegravir-based SBP, as discussed in the text

**Reference compound: 2-[6-methyl-4-(1-methyleneallyl)-2-oxo-1H-quinolin-3-yl]acetic acid, a commercially available close-in analog of 2-(6-chloro-2-oxo-4-phenyl-1H-quinolin-3-yl)acetic acid, compound 3 from Ref 76.

***Reference compound: compound 1 from Ref 76.

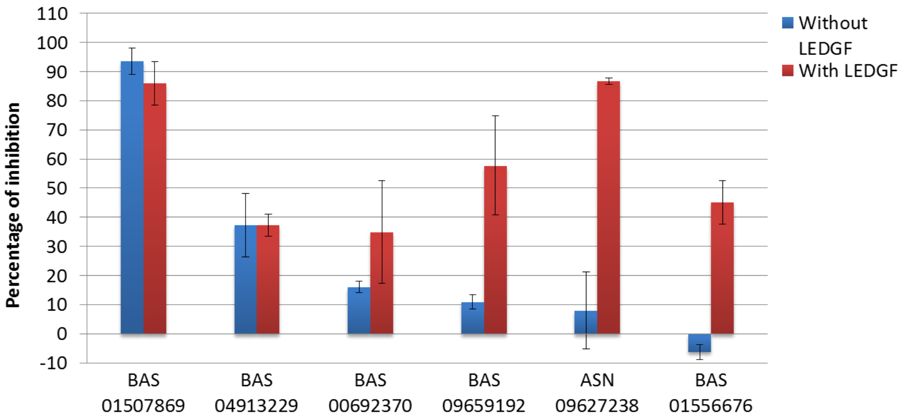


Figure 31: Percent inhibition of IN in the presence (red) and absence (blue) of the LEDGF activator.

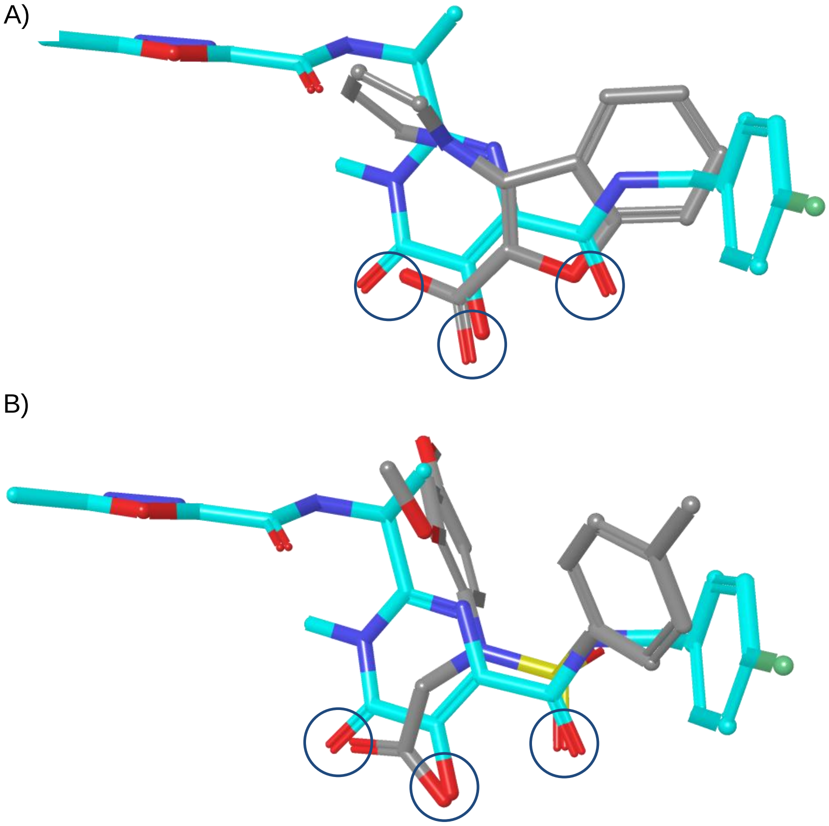


Figure 32: Two examples for hits matching the Raltegravir SBP. Hits (grey carbons) are displayed superimposed onto Raltegravir (cyan carbons). The Raltegavir metal chelating atoms and their matching

atoms from the hit compounds are circled. A) Asinex BAS 01507869 (94% inhibition, Table 4). B) Asinex BAS 02179070 (67% inhibition, Table 4).

4.C.2. Discussion

A structure-based screening protocol has been developed and applied to the discovery of novel inhibitors of HIV-1 binding to its cellular activator LEDGF. 17 out of 45 tested compounds were found to be more active than a commercially available reference compounds with an expected IC_{50} of approximately 12μM, ranging from 62%-100%. This corresponds to a remarkable 38% hit rate. Ten of the active compounds showed inhibition > 75%, with three compounds showing 100% inhibition and three over 90%. In comparison, reference compound activity was less than or equal to 60%. In addition, several of the less active compounds were found to be very potent in the presence of LEDGF, suggesting that while their direct inhibition potency is insufficient, they still efficiently displace LEDGF, which is the most biologically relevant function. Subsequent computational analysis suggests that some, but not all, may also be active-site inhibitors and are therefore potentially dual-acting compounds. Following the completion of this work, a compound specifically designed to have such dual activity was reported by De Luca et al.[89]. All hit compounds are chemically distinct from previously reported inhibitors of the IN/LEDGF interaction, including the recently reported De Luca compound, sharing only the pharmacophoric similarity integrated into the screening procedure. Interestingly, some of the hit compounds present an uncharged pharmacophore, which has not been described before[i]. Two out of the four tested uncharged compounds were active, corresponding to an exceptional 50% hit rate for this novel pharmacophore. Another interesting finding was that one compound potentiated IN activity. This novel activity still requires validation and its potential mechanism is yet unclear.

My original hypothesis (Project 2 discussion) suggested that compounds designed to inhibit LEDGF binding to IN may also function as allosteric inhibitors on IN enzymatic activity, similarly to the LEDGF-derived peptides. This provided justification for using a direct enzymatic inhibition assay or the initial *in-vitro* screening. Surprisingly, subsequent computational analysis revealed that several compounds from the Christ set as well as a number of the current hits may also be active site inhibitors, which is a more direct interpretation for the current assay but was unanticipated since there is no reported similarity between the LEDGF binding site and the IN active site. It is intriguing that these seemingly unrelated binding sites appear to share pharmacophore requirements.

---

[i] While this thesis was under review, natural product inhibitors of the IN-LEDGF interaction have been reported by Hu et al. (Hu, G.; Li, X.; Zhang, X; Li, Y; Ma, L.; Yang, L.M.; Liu, G.; Li, W.; Huang, J.; Shen, X.; Hu, L.; Zheng, Y.T.; Tang, Y.; J Med Chem 2012, 55, 10108−117). These include a set of chalcones, oleanolic acid derivatives, rosmarinic acid and 4-(benzyloxy)-2-hydroxy-N-(3-nitrobenzyl)benzamide, which all displayed submicromolar to low micromolar activity in an AlphaScreen binding assay. While none of these inhibitors resembles the ones we discovered, the chalcones are similar to our uncharged compounds in forming two H-bonds instead of the carboxylic-acid salt-bridge.

Measured activities should now be validated with dose response analysis in the presence and absence of LEDGF to assess both types of activities. Simultaneous activity at distinct sites could then be validated by curve shift assays assessing direct competition between the small molecule inhibitors and LEDGF or LEDGF-derived peptides.

Finally, I suggest that dual acting compounds may be especially interesting from a pharmaceutical perspective as they could be more difficult to overcome with escape mutations.

4.D. Project 4:

4.D.1. Results

CFTR was modeled using a unique combination of homology modeling and ROSETTA refinement. In the first modeling stage, initial outward facing models were generated by homology modeling based on the Sav1866 template. Given the low sequence similarity at the TM domains between CFTR and the available structural templates (<14% identity in the TMDs), detection and pairwise alignment of TM helices was performed by consensus analysis of several prediction methodologies and combined with evolutionary conservation and hydrophobicity analysis (Methods and Appendix B).

In the second modeling stage, carefully designed constrained ROSETTA refinement was applied to three of the best outward facing structures, effectively transforming them into more "channel-like" compact conformations while refraining from collapsing their pores (Methods, Figure 33). The latter was achieved by the continued inclusion of respective chloride columns during the refinement process. A total of 1,200 different model structures were generated at this stage.

Thirty models with the most favorable channel-like structures were subsequently subjected to rigorous analysis according to the following criteria: 1) quality of experimentally suggested salt bridge interactions (Appendix B Table S1) and salt bridges restrained during modeling (Methods); 2) agreement of channel pore dimensions calculated by HOLE2[90] with experimental data; 3) degree of consistency with all pairwise distances derived from experimental cross-linking data (Appendix B Table S2); 4) agreement with experimental residue accessibility data (Appendix B Table S3).

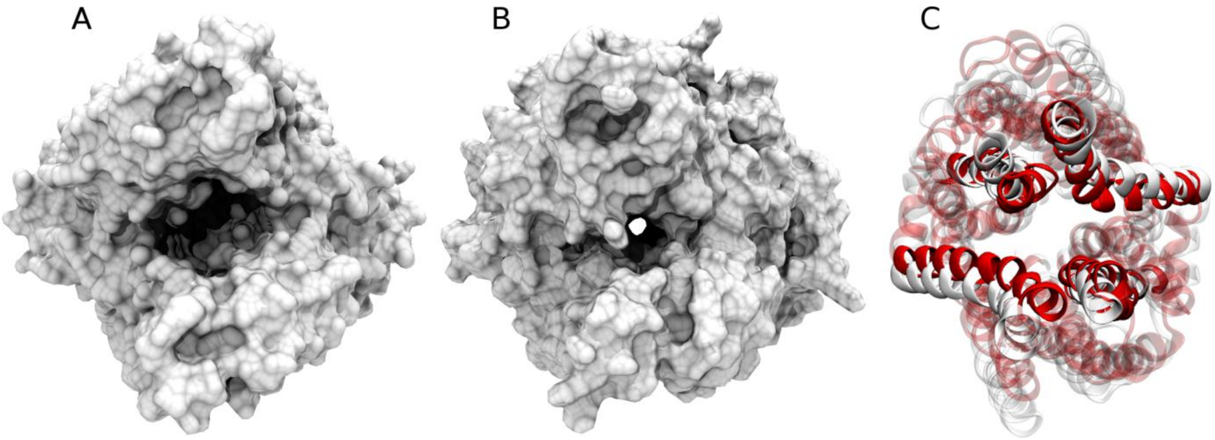


Figure 33. (A) Outward-facing conformation of Sav1866 compared against the more compact "channel-like" conformation of a representative CFTR model, as viewed from the outer membrane side, visualized with VMD[91]. (A) The molecular surface of Sav1866 outward-facing structure (2HYD). (B) The molecular surface of one of the ROSETTA models that passed validation, showing a more defined, compact and continuous pore. (C) A cartoon representation of a transmembrane-region superposition of the same ROSETTA CFTR model (in red) and Sav1866 (in grey). The four central helices of CFTR: TM6, TM1, TM12 and TM7 (anti-clockwise from top-left respectively, and in opaque red) are observed to move inwards, compared to the corresponding helices of Sav1866 (in opaque gray).

The three best channel-like structures surviving these filters were further scrutinized by molecular docking. Five previously identified anionic open channel blockers have been suggested experimentally to block the channel from the intracellular side, forming ionic interactions with K95[50]. Two additional anionic blockers have been found to block from the intracellular side but their ability to block K95 mutants has not been tested[92]. These seven compounds (Figure 34) were docked into the channel pore of the three alternative models using Glide[83]. Of the three models, one was found to out-perform the other two in terms of producing experimentally-supported blocker binding modes. Specifically, in this top performing model, highly overlapping pore-blocking poses were found for all seven blockers in which ionic interactions are formed with K95.

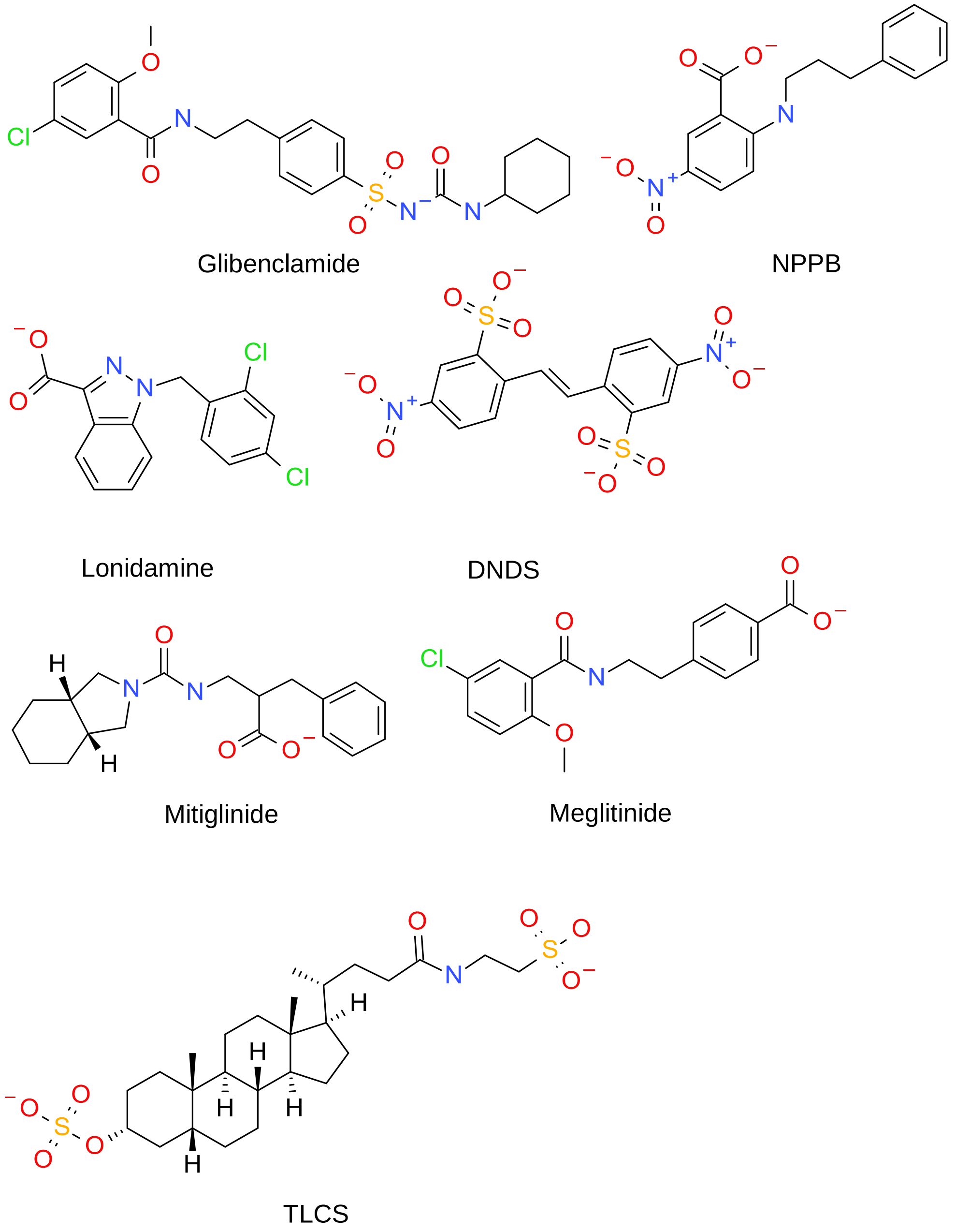


Figure 34: Structures of open channel blockers.

4.D.1.A. Construction of an "integrated model"

The three surviving models showed different strengths and weaknesses when evaluated by the above criteria. Realizing that the conformational sampling procedure described above is not ergodic and depends on initial conditions, an integrated model was constructed by combining the strong points of the two top scoring models. The third model was discarded as it was systematically out-performed by the other two.

This was accomplished using multi-template homology modeling with Prime[66], with one model providing the template for the TM11, TM10 and the interconnecting loop (better agreement with F508 related distance constraints) and the other providing a template for the rest of the structure (superior blocker docking results). Using Prime ensured that the respective side-chain orientations from each of the two templates were preserved in the combined model. Prime side chain refinement was subsequently performed on the grafted TM10-TM11 and proximal (3Å) residues to relieve unfavorable contacts at the seam line between template structures.

4.D.1.B. Evaluation of the integrated model

The final integrated model was re-evaluated using the above criteria. In this model, the expected channel topology is clearly visible (Figure 35) with the narrowest region about 4.5Å in diameter. The outer vestibule is dominated by R334 as previously suggested by Smith et al.[52] and by Linsdell[50,51]. Narrowing starts cytosolically to T338 (TM6) in agreement with the results of Alexander et al.[93], progresses through the selectivity conferring S341 (TM6), and culminates at M348 (TM6). The narrowest region of the modeled pore stretches down to W1145 (TM12) and is dominated by TM6 residues I334, V345 and M348, which were found to be of limited accessibility to thiol-reactive probes[93], TM12 residues S1141 and M1137, which were suggested as pore-lining[94,95], Ser877 (TM7) and W1145 (TM12) for which no functional data has been published so far, and the structural R352 forming a salt bridge with D993[96]. Widening of the inner vestibule begins cytosolically to W1145.

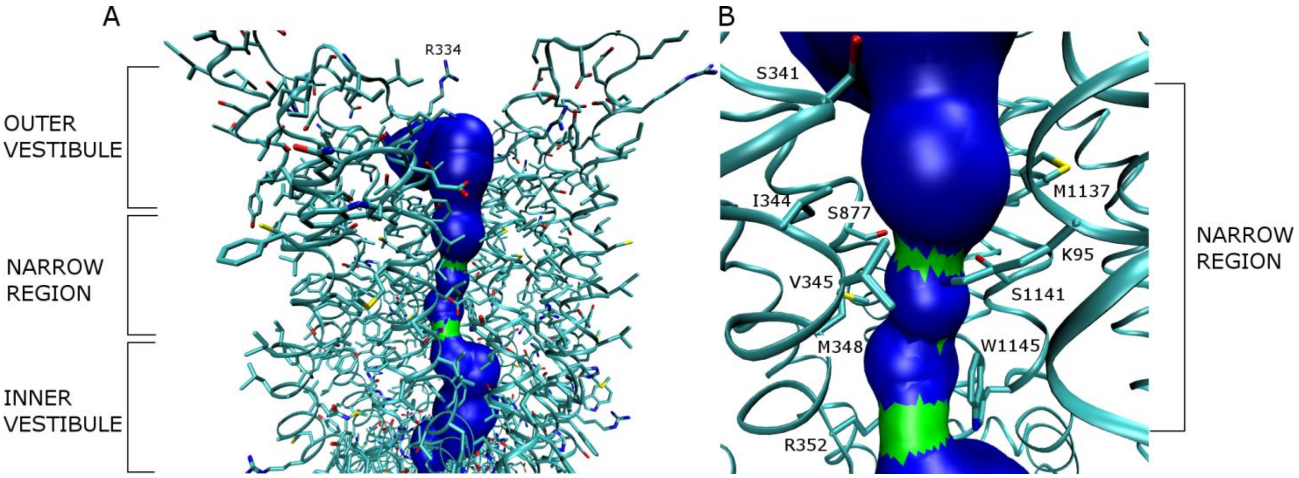


Figure 35. The modeled pore of CFTR visualized with VMD. A) The transmembrane region is divided into outer and inner vestibules separated by a narrow region. The protein is shown in cyan (as ribbons and side-chains) and the predicted pore (calculated with HOLE2) is shown in blue (pore diameter > 4.6 Å) and green (pore diameter < 4.6 Å). B) The narrow region starts cytosolically to T338 at approximately S341 and reaches W1145 at the intracellular side. A minimal pore diameter of ~4.5 Å is obtained between residues K95, V345, S877 and S1141.

4.D.1.B.1. TM6 Accessibility

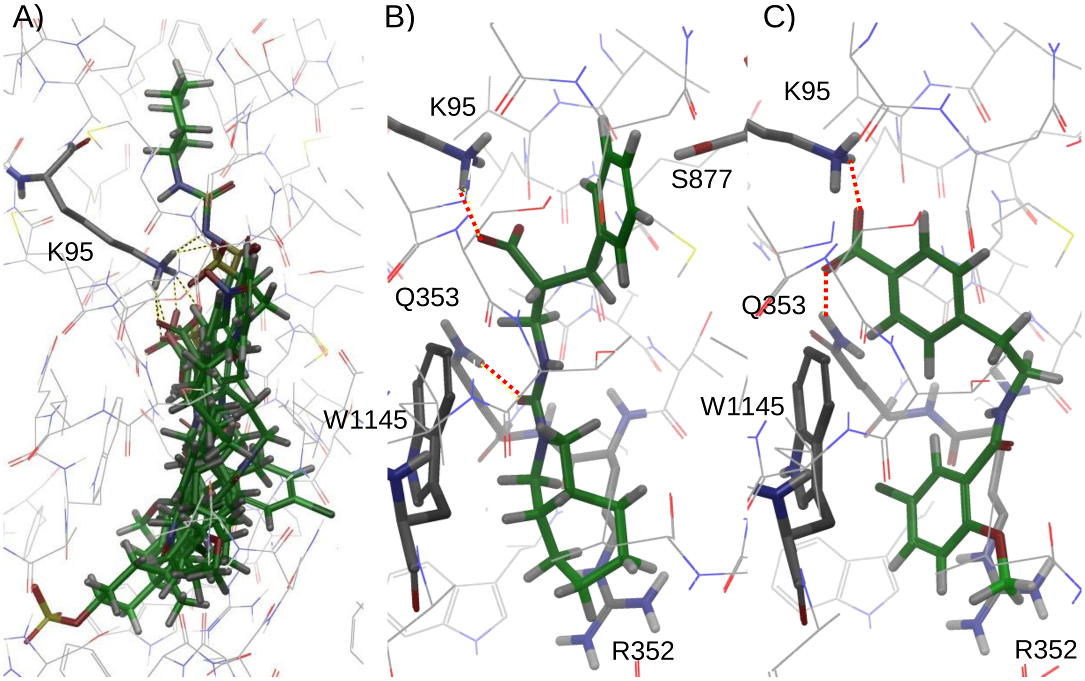
Due to the important contribution of TM6 to pore structure and function, a solvent accessibility assessment of TM6 residues was performed. Twenty four TM6 residues were individually mutated to Cysteine in Maestro followed by NACCESS[97] calculation on each mutant. The accessibility of the SG atom was then compared with experimental results describing the reactivity of each Cysteine with thiol reactive probes. Setting a cutoff of $20Å^2$ to predict experimental accessibility based on the NACCESS calculation, results in good predictive power with true positive rate = 0.6 and true negative rate = 1 (Table 5). This prediction scheme classifies nine out of the 11 positions of limited accessibility as inaccessible. Since these positions can only be modified by small permeant probes this can still be viewed as providing a good estimate of channel dimensions.

Table 5: Atom accessibility of mutant Cysteine SG atoms. Reactive: reactive to covalent modification by MTS reagents. Limited: reactive only to small permeant probes$[Ag(CN)_2]$- and $[Au(CN)_2]$-. Unreactive: does not react with any of the tested probes. Predicted reactivity is based on a 20Å^2 atom accessibility cutoff. These predictions correspond to a true positive rate = 0.6 and true negative rate = 1.

Residue Number	*Experimental data (Appendix B Table S3)*	*Atom Accessibility [Å^2]*	*Predicted Reactivity*
330	**Unreactive**	15.011	*Unreactive*
331	Limited	19.462	*Unreactive*
332	**Unreactive**	18.54	*Unreactive*
333	Limited	10.769	*Unreactive*
334	**Reactive**	57.398	*Reactive*
335	**Reactive**	40.589	*Reactive*
336	Limited	0.784	*Unreactive*
337	Limited	29.998	*Reactive*
338	**Reactive**	45.131	*Reactive*
339	Limited	0	*Unreactive*
340	Limited	4.032	*Unreactive*
341	Limited	18.408	*Unreactive*
342	Limited	18.888	*Unreactive*
343	**Unreactive**	0	*Unreactive*
344	Limited	5.034	*Unreactive*
345	Limited	9.928	*Unreactive*
346	**Unreactive**	7.158	*Unreactive*
348	Limited	17.349	*Unreactive*
349	Limited	2.109	*Unreactive*
350	Unreactive	0	*Unreactive*
351	**Unreactive**	10.368	*Unreactive*
353	Limited	43.614	*Reactive*

4.D.1.B.2. Binding modes of open channel blockers

When docking the seven open channel blockers to the integrated model, all adopt highly overlapping pore blocking binding modes, forming ionic interactions with K95 (Figure 36). This is in excellent agreement with the experimentally proposed mode of interaction for Glibenclamide, NPPB, Lonidamine, DNDS and TLCS. In addition, this provides a functional prediction for Mitiglinide and Meglitidine, which are known to block from the intracellular side but have not yet been tested for interaction with specific amino acids[92]. All of the docked open channel blockers also share an aromatic/hydrophobic interaction with W1145 and a Cation-π/hydrophobic interaction with R352. While R352 has an important structural role, no functional data is yet available for W1145. Additional residues implicated in individual blocker binding are S877, S1141 and Q353.



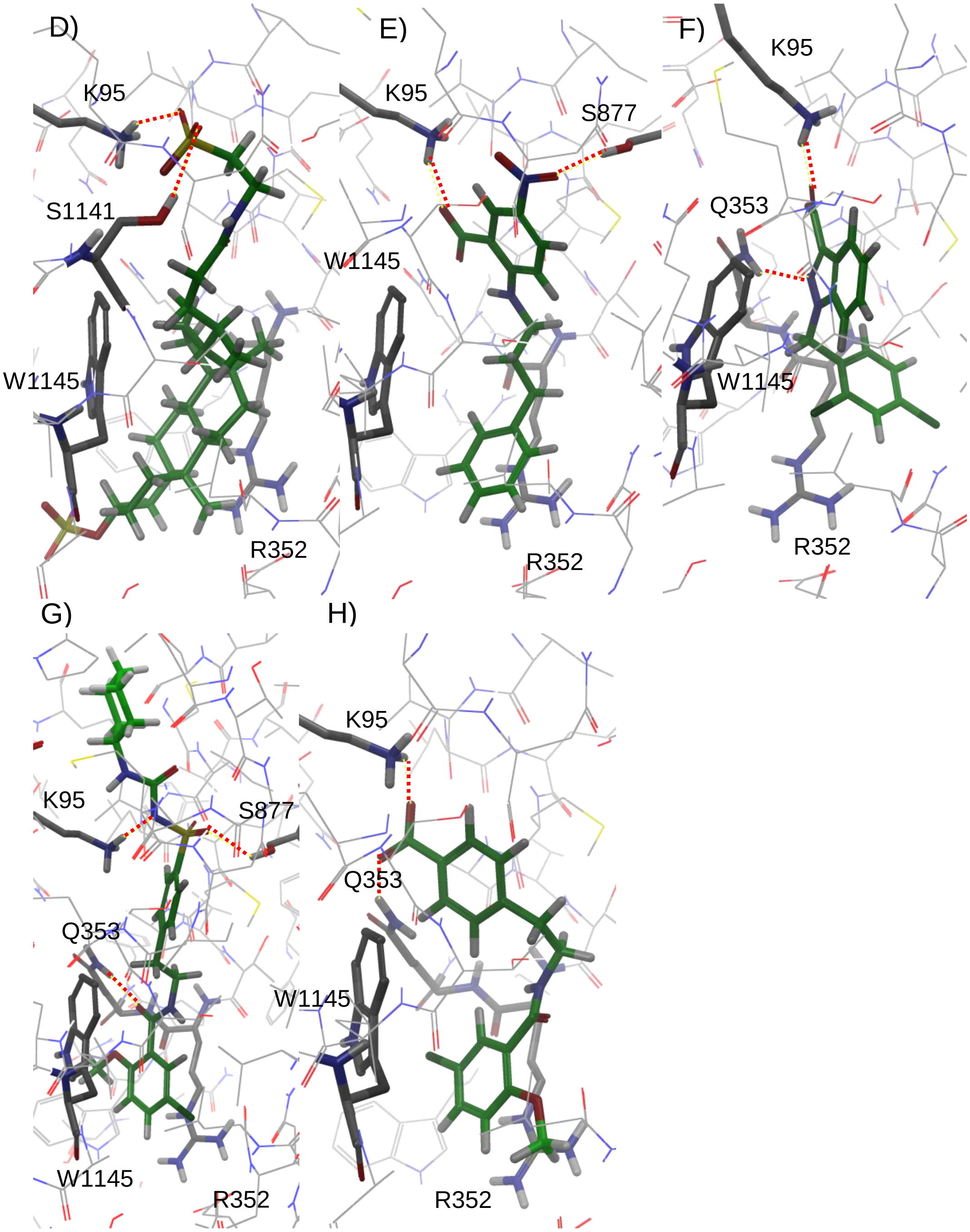


Figure 36: The binding modes of open channel blockers: A) Open channel blockers superimposed; B) Mitiglinide; C) DNDS; D) TLCS; E) NPPB; F) Lonidamine; G) Glibenclamide; H) Meglitidine. Common interactions are observed with K95, W1145 and R352. Dashed lines represent hydrogen bond interactions.

4.D.1.B.3. Evolutionary information

The final model was also evaluated and compared to previously published models using ConQuass, a novel quality assessment program[98], which is based on the consistency between the model structure and the protein's conservation pattern. The expectation is that, on average, the protein core will be much more evolutionarily conserved than the periphery. Validation of ConQuass on CASP8 results suggests that structures assigned ConQuass scores under -0.02 are mostly of limited quality, while ConQuass scores over 0.12 mostly correspond to good quality models[98]. The ConQuass score obtained for the current model was 0.19, exceeding the 0.12 cutoff, while lower scores are obtained for the previously published models (0.12 for Mornon and 0.07 for Serohijos).

4.D.1.B.4. Virtual screening

As a final validation, a virtual screening (VS) enrichment experiment was performed in order to test the model's ability to discriminate known open channel blockers from a set of random compounds. A random collection of 1,643 drug-like compounds, all possessing at least one negatively charged atom (common to known blockers and essential for interaction with K95) was extracted from the Maybridge screening collection (www.maybridge.com). Compound preparation, filtering and docking performed as described in Methods. A constrained docking approach was used for structure-based screening in which a single pore-blocking pose was extracted for each docked compound using constraints incorporated into Glide and applied post-docking. These constraints included a requirement for ionic interaction with K95 and a set of 3Å spherical positional constraints defined along the Cl^- column (Appendix B). The approach was first applied to the seven known blockers and reported poses were visually inspected in the channel pore. The process proved capable of extracting the best pose obtained in the unconstrained docking run and was therefore considered suitable for application in a virtual screening experiment.

In addition to the seven known blockers, 1,438 compounds out of the 1,643 survived the pose filter. Significant enrichment was subsequently obtained using Glide GScore as well as a score measuring the Coulombic interaction with K95 (also calculated by Glide) (Figure 37). An even better enrichment was obtained with a consensus scoring approach, applying cutoffs based on the scores obtained for the known blockers: 1) Glide Score < -5.0 kcal/mol; 2) Coulombic interaction energy with K95 < -44 kj/mol. Merely 226 library compounds (13.7%) survived this filter, including all known blockers, corresponding to an excellent $EF_{15\%}$ of 7.3, i.e. 100% knowns found within 13.7% of the screened library. This suggests the model has significant discriminatory ability in selecting for active compounds.

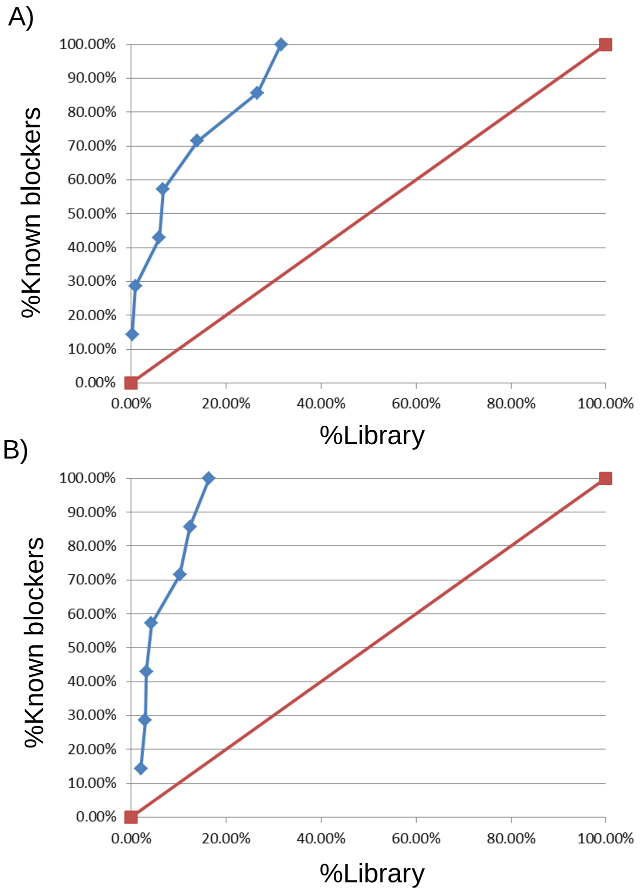


Figure 37: Enrichment graphs. A) Enrichment by Glide GScore; B) Enrichment by K95 Coulombic interaction; Blue: enrichment of known blocker. Red: random selection.

4.D.1.C. Modeling the F508del mutant

In order to investigate the structural effect of the F508 deletion, a model of the F508del mutant protein was generated from the *wt* by re-modeling with an alternative NBD1 template: a mutated structure missing F508 (PDB code 2PZF[99]) (Methods). Both *wt* and mutant models were subjected to molecular dynamics (MD) simulations in explicit, hydrated, lipid bilayer, with restrained chloride columns in their respective pores. Since the number of chlorides that pass simultaneously through the CFTR pore is not known, I assumed that the ions pass through the channel one at a time, and modeled the chloride column as having a single ionic charge distributed along the 20 comprising ions (Methods). Analysis of model stability during the simulations revealed that the local perturbation caused by the F508 deletion causes the mutant model to accumulate a significantly larger structural deviation from the initial "conducting state" conformation compared to the *wt* (Figure 38A). This may indicate a conformational instability of the "conducting state" which would correlate with the experimental finding that the few mutant channels reaching the membrane suffer from thermal instability[45,100,101]. This result further supports the validity of the current model. Interestingly, the effect spreads beyond the limits of the NBD1-ICL4 interface and throughout the protein structure (Figure 38B).

In the original F508del model, residues 491-513 were based on the 2PZF template. In order to check the sensitivity of the results to the extent of perturbation caused at the NBD1-ICL4 interface by incorporation of the mutant NBD1 segment, another model was constructed with a much smaller perturbation. In this model, a minimal perturbation was introduced by incorporating only residues 506-513 from 2PZF. This model was subjected to similar MD analysis, also revealing reduced stability of the mutant although of lower magnitude than that observed for the larger perturbation model (Figure 38A).

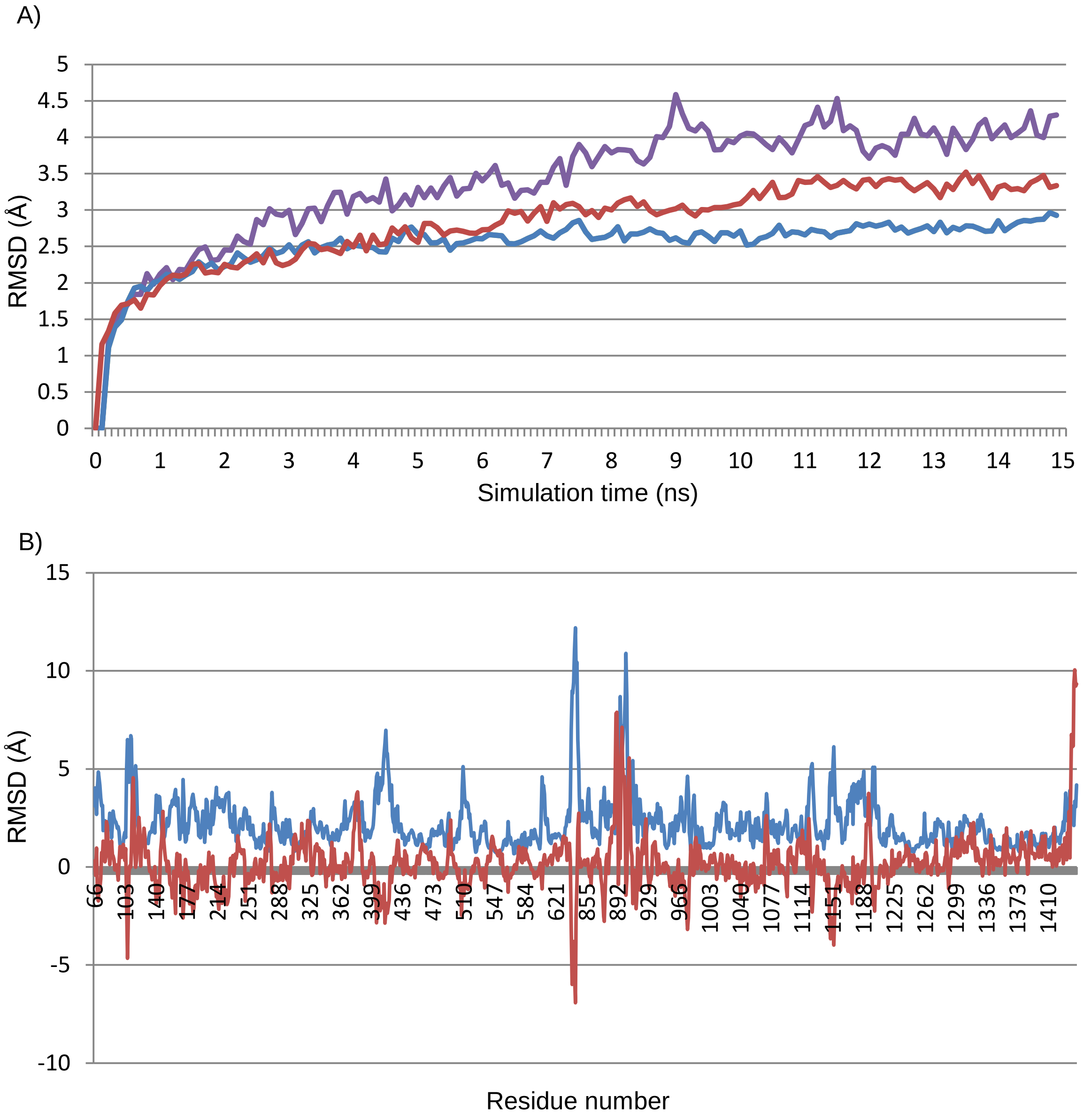


Figure 38: During MD, the F508del model deviates more than *wt* from the initial "conducting state" conformation. A) CA RMSD from the initial relaxed structure. Blue: *wt*; Red: Mutant (506-513 from 2PZF), small perturbation; Purple: Mutant, larger perturbation (491-513 from 2PZF). The larger the perturbation to the NBD1:ICL4 interface the larger the overall deviation from the initial "conducting state" conformation. B) Per-residue RMSD from the initial relaxed structure calculated over the 15ns MD trajectory. Blue: *wt* model. Red: $\mathrm{RMSD}_{\mathrm{F508del(506\text{-}513\ from\ 2PZF)}}$-$\mathrm{RMSD}_{wt}$. Per-residue RMS deviations of the F508del model are larger along most of the structure, despite the apparently local perturbation.

4.D.2. Discussion

My model of *wt* CFTR agrees with expected properties of the conducting state conformation. It has a channel-like structure with distinct outer-vestibule followed by a narrow region around the selectivity filter and an inner vestibule wide enough to accommodate open channel blockers. This is in contrast to previously published models presenting CFTR in the outward-facing Sav1866-based conformation corresponding to a wide barrel-shaped channel pore. A possible exception is a model recently reported by Alexander et al.[93], which was generated by MD refinement of a Sav1866-based outward-facing homology model. However, neither the sequence alignment to Sav1866 nor the model coordinates were made available to the public, preventing proper evaluation.

The present model provides insight into existing experimental data as well as novel predictions. From a structural perspective, the model and MD simulations propose structural and functional roles for several unaddressed residues. Especially interesting are five salt-bridge interactions suggested as important for the structural integrity of the conducting state: R352-D933, R347-D924, R134-E1104, R1102-D1154 and D873-R933. While experimental evidence exists only for the two, maintaining all five salt-bridges was found to have a positive effect on modeling outcome (data not shown) and all remained stable throughout unconstrained MD simulations. In support of a salt bridge between R134 and E1104, R134E/Q mutations were found detrimental to channel function[102]. Notably, both experimentally supported salt-bridge interactions are absent from the Serohijos *et al.* model, and in the Mornon *et al.* model only the R347-D924 salt-bridge exists (Appendix B table S1). The latter was reported as missing in the Alexander *et al.* model.

From a pharmacological perspective, my model reveals a common mechanism of inhibition by anionic open channel blockers, correlating with experimental results and predicting previously unaddressed interactions with pore residues. Furthermore, when tested in a virtual-screening enrichment experiment, the current model performed well, suggesting that it might be reliable enough to be used as a template for discovery of novel open channel blockers, which are recognized as potential antidiarrheal therapeutics. Molecules from the Maybridge collection that score highly in this experiment show excellent pharmacophore overlap with the known blockers, suggesting that they might indeed be active as open channel blockers. Reassuringly, without use of constraints imposing specific ligand chemistry, these compounds include chemical classes characteristic of CFTR channel blockers, such as disulfonic stilbenes, sulfonylureas, arylaminobenzoates, benzoic acid derivatives and 3-phenylpriopionic acid derivatives[92,103]. Examples for high-scoring library compounds are presented in Figure 39.

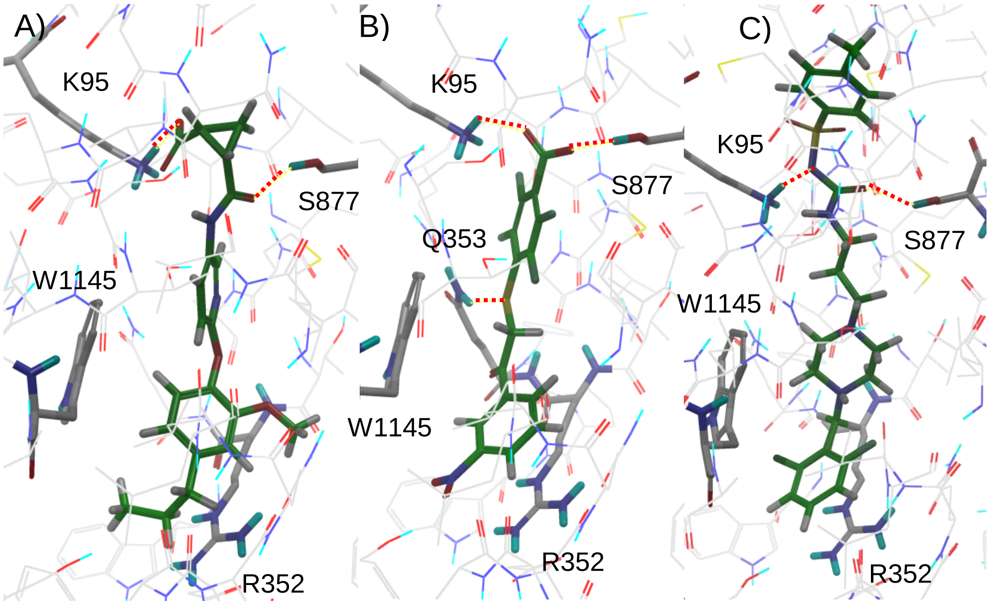


Figure 39: Examples for high-scoring library compounds. A) Maybridge BTB09406 (GScore = -7.3kcal/mol, K95 score = -62.3kj/mol); B) Maybridge BTB08329 (benzoic acid, GScore = -7.1kcal/mol, K95 score = -45.9kj/mol); C) Maybridge AW01120 (sulfonylurea, GScore = -7.0kcal/mol, K95 score = -44.2kj/mol). Dashed lines represent hydrogen bond interactions.

I also used the conducting state structure as a template for modeling and rationalizing the effect of the F508del mutation. In agreement with the known thermal instability of the mutant channels that manage to reach the plasma membrane and assume a functional form, the F508del model structure showed a significantly larger deviation from the initial “conducting state” conformation during MD simulations compared to the *wt* model. The surprising finding was that despite the seemingly small perturbation (Figure 40), the resulting structural changes were not limited to the NBD1-ICL4 interface but rather propagated across the entire structure. Since the F508 deletion has been implicated in reduction of NBD1 folding-yield and destabilization of interdomain contacts, it is difficult to predict the mutant conducting state conformation based on the *wt* model. Still, the results highlight the importance of the NBD1-ICL4 region for the structural integrity of the CFTR channel, where even a very small perturbation leads to significant structural deviation from the *wt* “conducting state” conformation and the larger the perturbation the greater the deviation as observed in Figure 38A.

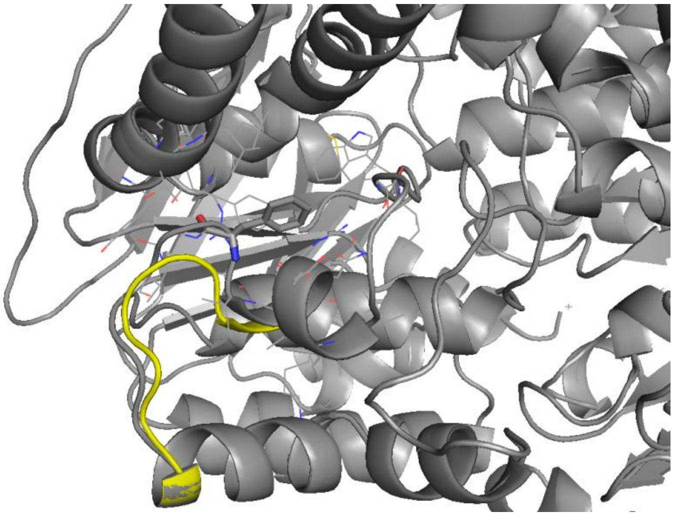


Figure 40: Perturbation caused by F508 deletion. Grey: *wt* model. Yellow: region of the F508del model based on residues 506-513 of the 2PZF X-ray structure of F508del NBD1. In this model of F508del CFTR there is no real perturbation to packing interactions except for the deletion of the F508 side chain. Still, structural destabilization is observed during MD simulations (Fig. 38).

On the down side, recent experimental results published after this work was completed expose a limitation of the current model and modeling approach. The recent publication suggests a cross-linking distance of 7-8 Å between I344 and both K95 and K98[104]. In contrast, inter-residue distances in the current model are 11.6 Å and 15.3 Å for K95-I344 and K98-I344, respectively, suggesting a slightly different relative orientation of TM1 and TM6. This reflects the inherent limitation of the current modeling procedure as well as other modeling protocols attempting to derive the conducting state of CFTR from the Sav1866-based outward facing conformation (e.g. using MD as mentioned above). On the one hand, the Sav1866 based conformation is not necessarily relevant for CFTR. On the other hand, the template TM region is inherently flexible, having to shift between inward-facing and outward-facing conformations. The combination of these two issues suggests that obtaining conformational convergence would be hard and could rely heavily on modeling conditions, e.g. the experimental constraints incorporated into the modeling procedure. While I have obtained a conformation that is in much better agreement with experimental data than previously reported models, it may not be the most accurate conformation or alternatively, it may be one of several conformations accessible to functional CFTR: a concept that is supported by recent experiments[105]. I expect that as more structurally interpretable experimental data referring to the conducting state becomes available the current modeling method could be used to derive models of increased accuracy.

The current modeling method is also limited by the approximations inherent to the use of classical ROSETTA which assumes solvation in water and I realize that the results may improve if a newer version of ROSETTA capable of handling a membrane environment is used. Still, this may not have had a critical effect on modeling outcome given the constrained nature of the refinement and in light of the clear agreement with a large bulk of experimental data.

Finally, because the outward facing Sav1866-based conformation may not be realistic for CFTR, it is unclear whether physically realistic and computationally demanding MD simulations, such as those used by Alexander et al., are a good algorithmic choice. The present modeling approach, which is less CPU intensive, essentially provides an experimentally guided conformational search mechanism which could be followed by full blown MD refinement in an explicit lipid bilayer once the structure is much closer to its target conformation. This might be a more adequate approach in this particular case.

5. Methods

5.A. Project 1:

5.A.1. Modeling protein-ligand complexes

1YCR-based protein-ligand complexes were generated by direct modification of the MDM2-p53(17-29) co-crystal structure using the Schrodinger Maestro Build module[106]. Ser15 and Gln16, missing from the crystal structure, were added to all relevant peptides in extended conformation. 1T4F-based complexes were generated by superimposing peptides from 1YCR-based complexes onto the 9-mer ligand in 1T4F. The p53(15-29)Pro27Ser mutant was modeled in both *wt* and helical conformation. The helical conformation was generated from *wt* by imposing a helical structure on the p53(24-29) residues using the Accelrys Discovery Studio 2.0 Protein Modeling tool[107]. Residues 24-29 were selected to maximize the overlap between *wt* and helical conformations. Peptidomimetics compounds 3,4,5 and 8 from the Garcia-Echeverria et al. paper[63] were modeled based on the crystal structure of compound 8 with MDM2 (pdb code 2GV2[108]). In the 2GV2 crystal, the peptide lacks the ACE and NH2 termini described in the Garcia-Echeverria paper. Thus, termini were added in Maestro[106]. Beta-hairpin compounds 62, 64, 67, 77a and 78a from Fasan et al.[64] were modeled based on the crystal structure of compound 78a with MDM2 (2AXI[64]). These compounds were chosen to span a range of activities.

Complexes were prepared for simulation with the Schrodinger protein preparation tool[106]. Following addition of hydrogen atoms, protein-preparation optimizes the orientation of hydroxyl and sulfhydril groups, the protonation and tautomeric states of His side chains and the "flip" state of Asn, Gln and His side-chains around their outer most Chi angle. Then, a series of short restrained minimizations is performed, optimizing the modified structure while keeping it as close as possible to the crystal structure coordinates. The first minimization iteration reorients side-chain hydroxyls and sulfhydryls by tightly tethering all non-hydrogen atoms with a force constant of 10 kcal/mol·Å^2 and minimizing with torsion interactions turned off. The following minimizations use a restored torsion potential and progressively weaker position restraints on the non-hydrogen atoms, using force constants of 3, 1, 0.3 and 0.1 kcal/ mol·Å^2. Each minimization terminates at 100 steps, when the potential energy gradient reaches 0.05 kcal/mol·Å or when the heavy-atom RMSD deviation from the initial coordinates reaches 0.3Å. Conjugate-gradients minimization is performed using *Impact*[109] with the OPLS_2001 force-field and a distance dependent dielectric of 4.

5.A.2. Activity data for peptidomimetics

Comparing activities of compounds studied in different types of assays and/or assay conditions requires Kd rather than IC_{50}. Kd values were obtained for the beta-hairpins by converting the reported IC_{50} values using the Cheng-Prussof equation[110]. The data required for conversion were not available for the 8-mer compounds. Thus, 8-mers were analyzed separately, while the beta-hairpins were analyzed both independently and in combination with the p53(17-26) analogues.

5.A.3. MacroModel single-point MM-GB/SA

Complexes were typed with the OPLS-2005 force-field. Solvent effects were modeled using a GB/SA implicit solvent model. Structures were minimized using Truncated-Newton minimization terminating after 2000 steps or after reaching a gradient of 0.01 kcal/mol. Following minimization, MM-GB/SA was calculated as $\Delta G = G_{complex} - G_{ligand} - G_{protein}$ using a set of perl and MacroModel python scripts.

5.A.4. CHARMM single-point MM-GB/SA

Complexes prepared using the procedure described above were typed using the CHARMM Momany-Rone force-field[107]. Solvent effects were modeled with the GB/SW implicit-solvent model. Structures were energy minimized using the Steepest-Descent algorithm followed by Adopted-Basis Newton-Raphson, both terminating after 2000 steps or after reaching a gradient of 0.01 kcal/mol. Following minimization, protein and ligand structures were extracted from the complex and MM-GB/SA was calculated as $\Delta G = G_{complex} - G_{ligand} - G_{protein}$ using a set of perl and CHARMM scripts.

5.A.5. MacroModel implicit solvent MM-GB/SA

Complexes were typed with the OPLS-2005 force-field. Solvent effects were modeled using a GB/SA implicit solvent model. Structures were minimized using Truncated-Newton minimization terminating after 2000 steps or after reaching a gradient of 0.01 kcal/mol. Following minimization, Stochastic Dynamics was performed with the following protocol: Gradual heatup from 100K to 300K over a period of 20ps;

Equilibration + production over a period of 3ns at constant temperature. All bonds to hydrogen atoms were constrained using the SHAKE algorithm allowing a timestep of 1.5fs. Coordinates were saved at 2ps intervals (Also see discussion in Supporting Information).

MM-GB/SA calculation: Single-trajectory MM-GB/SA was performed due to the high flexibility of the unbound 15-mer which may lead to inadequate conformational coverage. This method, in which the conformational ensembles of the unbound protein and the free ligand are extracted from the ensemble of the protein-ligand complex, has been shown to reach better convergence than the multiple trajectory approach[19,22,58]. MM-GB/SA energy terms (see Figure 3) were calculated from 100 snapshots sampled between 200ps and 400ps of the simulation (see Supporting Information for details) using a combination of perl and MacroModel python scripts.

5.A.6. Translating MM-GB/SA energies to pK_d values

MM-GB/SA predictions for eight p53(15-29) analogues correlated well with experimental affinities (R^2=0.77). The regression line describing the correlation has the equation: MM-GB/SA = -70.28*pKd + 180.84. This equation can be reorganized to describe pKd as a function of MM-GB/SA energy. The new equation reproduces pKd values for these eight compounds with an average unsigned error of 0.16 log units. When predicting the affinity of peptides without experimentally measured affinities, MM-GB/SA energies can be translated into estimated pKd values using the new equation and the prediction error can be roughly estimated using the calculated average error.

5.A.7. MacroModel simulations for assessing conformational stability

Peptides were typed with the OPLS-2005 force-field. Solvent effects were modeled using a GB/SA implicit solvent model. Non-bonded interactions were treated with a cutoff of 8Ǻ for vdW interactions and 20Ǻ for electrostatics, as required for MacroModel GB/SA simulations according to the program manual. Structures were minimized using Truncated-Newton minimization terminating after 2000 steps or after reaching a gradient of 0.01kcal/mol. The following MD cascade was then carried out: Gradual heatup from 100K to 300K over a period of 20ps; Equilibration + production over a period of 100ns at constant temperature. All bonds to hydrogen atoms were constrained using the SHAKE algorithm allowing a timestep of 1.5fs and Stochastic Dynamics was used for temperature control. Coordinates were saved at 50ps intervals. In the 1YCR co-crystal structure, p53 residues 19-23 are in helical conformation. Therefore, helical stability was estimated by superimposing all MD snapshots onto the initial structure using the backbone of p53(19-23) for calculating the transformation matrix and for deriving the RMS deviations.

5.B. Project 2:

5.B.1. Initial model construction

The basis for the tetramer assembly was the crystal symmetry observed in the 1ex4 X-ray structure of the HIV-1 IN CCD-CTD. Crystal mates passing within 5.0Å of the 1ex4 dimer were generated using the PyMOL[111] command “symexp sym,1EX4,(1EX4),5” and sym005 object was used as the scaffold onto which the other crystal structures was superimposed.

A full length model was generated by homology modeling using Prime. The different templates of each monomer were fused by homology modeling the IN sequence Q76353_9HIV1 (1-270) using the superimposed structures as templates. In the model tetramer monomer pairs A, B and C, D each correspond to a dimer in 1ex4 symmetry tetramer which is a dimer of dimers. To reduce clashes and optimize connectivity in the initial model monomers A and C were based on 3F9K (Q137-V150), 1EX4 (S195-D270), 1KY6 (K186-Y194, F1-K46), 2B4J (all the rest) and monomers B and D were based on 3F9K (F1-C43, V54-D55, Q137-V150, K186-Y194), 1EX4 (S195-D270), 2B4J (all the rest).

LEDGF IBD domains were added to the tetramer based on the 2B4J structure.

5.B.2. Protein-protein docking

The proximal N-terminal domains were docked using the PatchDock webserver (bioinfo3d.cs.tau.ac.il/PatchDock). Since the proximal NTDs originate from the two unit cell monomers observed in 1KY6, which present slightly different conformations of IN(1-45), three non-redundant combinatorial docking options were generated (A to A, A to C, C to C). Dimers with the geometry required for fitting back into the tetramer model were only found within the top ten PatchDock solutions. Therefore, the top ten solutions from each docking run (A-A, A-C, C-C) were extracted and subjected to refinement with FireDock (bioinfo3d.cs.tau.ac.il/PatchDock). The top thirty solutions were visually inspected and seven structures were selected for further analysis.

5.B.3. ROSETTA refinement

Rosetta relaxation was performed in fast mode with default parameters.

Solutions were clustered by backbone RMSD at 4Å cutoff using the GROMACS 4[112] g_cluster utility.

5.B.4. Introducing the dimers back into the tetramer model

The isolated dimers of N-terminal domains were superimposed in Maestro[77] onto the proximal N-terminal domains in the initial model. Backbone superimposition was found to produce better overlaps than protein structure alignment.

5.B.5. Relieving initial steric violations

Models were subjected to MacroModel minimization of the proximal NTDs and 5Å surroundings using the PRCG minimizer with a derivative convergence criterion set to 1 kJ/Å-mol and a maximum of 5,000 steps using the OPLS_2005 force field and a distance-dependent dielectric of 4. The minimized models were optimized with Prime side chain refinement of the same region using default settings.

5.B.6. Homology modeling with MODELLER

The IN(228-235) loops which form a part of the tetramer interface were removed from monomers B and D to enable de-novo modeling and initial tetramer models were generated using the automodel class without MD refinement. Tetramer interfaces were subsequently optimized using the loopmodel class with fast MD refinement. Optimized regions included IN(46-55) from monomers A and C, IN(186-194) from monomers A and C, IN(228-235) from monomers B and D, IN(54-58) from monomers B and D, all residues in the interface between LEDGF and distal NTDs. Ten solutions were generated for each model and sorted by PDF score[70] and as well as DOPE score[71]. Selection by

DOPE score was found to produce models with better side chain interactions than models selected by PDF score.

5.B.7. Prime side chain refinement

The protein preparation workflow in Maestro[77] was used for preparing MODELLER results for Prime refinement. Subsequent Prime side chain refinement was performed using default settings.

5.B.8. GNM calculations

GNM calculations were performed using software kindly provided by Prof. Türkan Haliloğlu (Bogazici University, Bebek, Istanbul, Turkey) and analysis of GNM results was performed using in-house programs.

<u>Analysis of differences in the mean square fluctuations</u>: mean square amino-acid fluctuations produced by the GNM program for the unbound structure were subtracted from those of the bound structure. The magnitude of the change was expressed as percentage of the fluctuations observed in the unbound structure.

<u>Analysis of differences in cross-correlations</u>: the extent of change in the cross-correlation of each residue with all other residues was estimated as the root mean square deviation between vectors of pairwise cross-correlation coefficients extracted for each amino acid from the matrices of cross-correlation coefficients produced by the GNM program: $RMSD_i = \sqrt{\sum_{j=1}^{n} \left(\rho_{i,j}^{u} - \rho_{i,j}^{b}\right)^2 / n}$, where n is the total number of residues in the unbound tetramer and $\rho_{i,j}^{u}$, $\rho_{i,j}^{b}$ are the correlation coefficients between residues i,j in the unbound and bound structures, respectively.

Visualization of cross-correlation differences: RMSD vectors to be compared were normalized to (-1,1) using the equation $RMSD_i^{norm} = \frac{\mathrm{RMSD}_i - 0.5(\mathrm{RMSD}_{min} + \mathrm{RMSD}_{max})}{0.5(\mathrm{RMSD}_{max} - \mathrm{RMSD}_{min})}$, where RMSD_{min} and RMSD_{max} are the minimal and maximal RMSD values in the vectors being compared. These were subsequently converted into a Pymol coloring scheme representative of the relative magnitudes of change.

5.C. Project 3:

5.C.1. Screening library preparation

Three libraries were selected for screening: Chembridge diversity libraries, DIVERSet and NOVACore/PHARMACophore (~50,000 and ~200,000 compounds respectively, www.chembridge.com); Asinex GOLD and PLATINUM collections (~600,000 compounds, www.asinex.com).

Library compounds were prepared using ligPrep[113], generating at most four tautomeric states with protonation assigned by Epik[114] at a pH range of 5.0 to 9.0, as recommended by the developers (personal communication). Prepared compounds were subsequently filtered in two steps. Initially, Maestro Ligand Filtering was used to discard all compounds not matching the following criteria: 1) Number of reactive groups = 0; 2) Number of hydrogen bond donors <= 5; 3) Number of hydrogen bond acceptors <= 10; 4) MW <= 600. The limit on the number of hydrogen bond acceptors and donors was taken from the Lipinski rule-of-five[115] and the MW limit was extended from Lipinski to encompass the larger size typical to protein-protein interaction inhibitors[116]. Subsequently, the program PPI-HitProfiler, based on recently derived decision trees[117], was used to carve out a final set of compounds with increased probability of containing protein-protein interaction inhibitors (the program was executed in soft mode corresponding to a higher sensitivity and lower specificity). This filtering process resulted in ~50% reduction in library size without any loss of known inhibitors.

5.C.2. Preprocessing of protein structures

In preparation for binding site analysis and later for docking, IN pdb structures were prepared as follows: 1) When the pdb entry contained a single monomer, the biological subunit was reconstructed using the Maestro “crystal mates” function; 2) All water molecules non biological ligands were removed; 3) Missing side-chain were completed using Prime; 4) preparation and optimization using the Maestro protein preparation tool[77] including addition of Hydrogen atoms, optimization of hydrogen bond networks and a final constrained energy minimization.

5.C.3. Structure-Based Pharmacophore (SBP) caclulations

5.C.3.A SBP construction for screening

The structure-based pharmacophore was constructed in Phase. Phase feature definitions were modified as follows: 1) the Aromatic feature definition was modified so that aromatic rings are mapped as hydrophobic features (achieved by unchecking the "Ignore" box for default_aromatic_surface). This modification causes the phenyl ring on cmp6 to produce a hydrophobic site in a position which is in good agreement with the SiteMap results; 2) the hydrogen bond acceptor feature was modified to include carboxylate oxygen atoms (achieved by adding a [O-X1] SMARTS pattern to the list of allowed acceptor patterns and unchecking the “Exclude” option from the two existing carboxylate entries). This enables Phase to produce two acceptor features on top of the compound 6 carboxylic acid oxygens, which would allow uncharged hydrogen bond acceptors to map as discussed in the Results. Receptor-based excluded volumes were added with a 2.0Å buffer.

5.C.3.B. Database construction for screening

A Phase database was generated for each screened library from the command line. The following parameters were modified from their default values: 1) amide bonds were retained in their original conformation generated by ligPrep; 2) the energy window was set to 25kcal/mol; 3) the maximum number of conformers per molecule was increased to 200.

5.C.3.C. SBP screening

The structure-based pharmacophore hypothesis was screened from the command line. The following parameters were modified from their default values: 1) mapping rules were modified to allow both hydrophobic and aromatic groups to match HR1 and HR2 and to allow hydrogen bond acceptor groups to match the AN feature, as discussed in Results 2) vector Weight was set to 0 to remove dependence on the carboxylic acid geometry when matching acceptor groups to the A1, A2 and AN features; 3) Prohibited matches were defined to prevent the following: mapping of Donor, Hydrophobic and Positive groups onto A1, A2 and AN, mapping of Acceptor, Donor, Negative and Positive features onto HR1 and HR2; 4) compounds were required to match at least 3 features by setting minSites to 3 (larger matches were still preferred); 5) deltaDist (alignment cutoff factor) was set to 1.5Å. 6) volumeCutoff was set to 0.2, imposing a loose requirement that matching compounds occupy a similar region of the biding site as compound 6.

5.C.3.D. SBP Post processing of docking results

Glide output files containing multiple binding modes for each docked compound were rigidly screened using Phase to identify binding modes matching the structure-based pharmacophore. The single, top scoring binding mode for each compound was extracted identified and extracted using the glide_sort utility.

5.C.3.E Raltegravir-based SBP

Structure of Raltegravir was extracted from 3OYA[118]. Phase hypothesis included three Acceptor features mapped onto the metal-chelating oxygen atoms. Negatively charged atoms were allowed to match these features and directionality was ignored in result filtration. Ten solutions were inspected for each mapped compound to evaluate alternative overlaps.

5.C.4. Glide Docking

A Glide5.6 grid was constructed around the co-crystallized compound 6 using default parameters. Glide SP docking was performed using default parameters except the following: 1) amide bonds were retained in their original conformation because all other options were found to produce cis-amide conformations (I reported this bug the software developers); 2) Epik state penalties were added to the Glide Docking Score to account for the prevalence of the docked tautomer, in accordance with recent recommendations by the Glide developers (http://www.schrodinger.com/newsletter/12/); 3) Five top scoring poses were written for each docked compound.

Calculations were distributed over a cluster of Quad-Core machines with AMD Opteron 2.2GHz CPUs and 16GB RAM.

5.C.5. Prime-MM-GBSA calculations

Prime-MM-GBSA calculations were performed with a 12Å flexible region.

5.C.6. Compound clustering

Hierarchical clustering was performed in Canvas using hashed linear fingerprints, the Tanimoto similarity metric and the average linkage method. Clusters were merged at a distance of 0.6 providing a good balance between number of clusters and compound similarity within each cluster.

5.C.7. Conformational Analysis

Conformational analysis was performed in MacroModel[68] with the Conformational Search Torsional sampling method with GB/SA water, 50 PRCG minimization steps and otherwise default parameters. For each analyzed compounds, the predicted solution ensemble was superimposed onto the docked conformation using the the Maestro Superposition option.

5.C.8. *In-Vitro* testing

Compounds will be tested at the lab of Prof. Assaf Friedler from the Hebrew University, in an assay previously described[119].

Prior to testing of purchased compounds, the assay will be calibrated using the following compounds: 1) compound 1, 2 from Christ et al.; 2) a compounds identical to compounds 3 except for a methyl instead of the chloro group in compound 3. Given the binding mode of compound 3 observed in 3LPT, this modification can be expected to have little effect on binding. A commercial source for these compounds was found by structure searches of the eMolecules collection (http://www.emolecules.com/).

5.D. Project 4:

5.D.1. Prediction of TM Helices and Pairwise Alignments

We approximated the location of the TM helices in the sequence CFTR using the TMHMM[120], HMMTOP[121], SPLIT4[122] and PSIPRED[123] algorithms. Since CFTR is distantly-related to sav1866, direct sequence alignment could not accurately match the two sequences. We therefore utilized three profile-to-profile fold recognition methods to obtain sequence alignments between the TM domains of sav1866 and those of CFTR, namely HMAP[124], HHPRED[125] and FFAS03[126].

5.D.2. Sequence Data and Evolutionary Conservation Analysis

The phylogenetic analysis by Jordan et al., identified that 11 other ABCC human proteins, exhibiting the same domain organization as CFTR, are CFTR's closest homologues in the human proteome[127]. This study showed that three human ABCB proteins, which reside on a separate branch on the phylogenetic tree, also encompass two TMDs and two NBDs on the same polypeptide chain. Relying on this analysis, we collected sequences detected by the Ensembl database[128] as orthologues for these 14 human ABCC and ABCB proteins. We then excluded fragments, aligned the remaining sequences with MAFFT[129] and constructed a phylogenetic tree using PHYML[130]. Using the phylogenetic tree, we observed that the proteins belonging to the ABCB subfamily were significantly distant from the ABCC subfamily. Therefore, we included in the final multiple sequence alignment (MSA) only the 177 sequences belonging to the ABCC subfamily. This MSA was used to compute evolutionary conservation scores via the ConSurf webserver[131] (http://consurf.tau.ac.il).

5.D.3. Homology modeling

An outward-facing conformation of CFTR was first modeled by homology to the Sav1866 transporter. Homology modeling of CFTR is a challenging task due to the low sequence similarity at the TM domains between CFTR and the available structural templates (<14% identity in the TMDs). Thus, detection and pairwise alignment of TM helices was performed by consensus analysis of several prediction methodologies. Core segments of TM helices were first identified by integrating the results of several profile-to-profile alignment methods, secondary structure prediction and TM predictions algorithms. Subsequently, data from evolutionary conservation and hydrophobicity analyses was used to fine-tune TM boundaries based on the following conventions: 1) residues comprising the protein core (e.g., inter-helical interfaces) of TM proteins tend to be conserved[132-135] while residues facing the membrane are in most cases less vital to protein function and are thus mostly variable; 2) TM helices should possess a strong hydrophobic signal to insert into the membrane, sequestering polar and charged residues from contact with the lipid tails. These guidelines were combined with functional data in order to reproduce experimentally suggested orientation of functional residues as best as possible (See Appendix B for details).

This structural bioinformatics analysis was performed in collaboration with Maya Schushan from the Ben-Tal lab.

The full-length *wt* CFTR was modeled by homology to Sav1866. Two crystal structures of Sav1866 are available: 2HYD, Sav1866 in complex with ADP, and 2ONJ, Sav1866 in complex with AMP-PMP, distinguished by slight conformational changes in the region of the bound nucleotide. AMP-PNP is a closer analog of ATP, making 2ONJ a more suitable template for modeling the CFTR conducting state, which requires bound ATP.

The initial homology model was built with MODELLER[70] using a multiple templating approach combining the Sav1866 structure and the structure of *wt* human NBD1 (2PZE).

The following restraints were used in modeling: 1) Distance restraints imposing the experimentally proposed salt-bridges: R352-D993[96], R347-D924[136]; 2) Distance restraints imposing salt-bridges inferred from structural analysis of initial models generated without these constraints: R134-E1104, R1102-D1154, D873-R933; 3) Helical constraint on the TM11 insertion required in order to preserve pairwise distances between F508 and neighboring residues as proposed by cross-linking experiments (Appendix B Table S2); 4) Helical constraint on TM8 in order to prevent unwinding due to the deletion required for reorienting D924; 5) Distance restraints between MG and ATP with the following residues: H1402, E1371, K1250, D1370 and S1251, comprising the ATP site of NBD2, thus ensuring they match their corresponding catalytic residues in Sav1866; 6) Distance restraints between pairs of residues that are known to associate in experimental cross-linking studies: 340-877[137], 348-1142[138], 351-1142[138], 95-1141[139], 356-1145[138], 508-1068[47], 508-1074[47].

A Cl^- column was inserted into the pore of the template structure of Sav1866, which was then made rigid during the modeling process. This was found necessary for preventing collapse of the channel pore during the next stages of refinement. The column was placed in line with the Z-axis, at X and Y equal to zero according to the 2ONJ coordinate system in the OPM database. Cl^- atoms were placed at a spacing equivalent to their vdW diameter, equating to a total of 36 Cl^- ions in the column.

Five crystallographic water molecules, identified by their conservation in crystal structures of NBD1 and other NBDs[99,140-144] as well as relatively low B-factors, were included in the NBD1 model. In addition, three of these water molecules were added in the NBD2 domain in the model.

The regulatory extension (RE) loop of NBD1 was modeled based on the previously generated model of Mornon et al.[48]. RE was excised from the Mornon model, superimposed onto 2PZE, and incorporated into the model by MODELLER multiple templating.

Ten models were generated with MODELLER with slow refinement and twenty cycles of simulated annealing. Of these, three were selected for further refinement, based on visual inspection of the salt-bridge residues and pore-lining residues.

Initial modeling was performed in collaboration with Dr. James Dalton. Modeler calculations were executed at TAU as well as at UPF.

5.D.4. ROSETTA refinement

The TM and NBD2 regions of three different models were refined using the ROSETTA v3.1[69] 'Fast Relax' refinement procedure, while keeping the NBD1 domain frozen and employing rotamer-based salt-bridge restraints similar to those used in model building. (Initial testing, showing that this procedure tends to generate a collapsed channel, led to the inclusion of a Cl^- column positioned manually at the center of the channel pore). A total of 1,200 ROSETTA conformations were generated (400 for each MODELLER output structure), and subjected to filtering based on salt-bridge interactions. Merely 170 models passed this filtering stage, all belonging to the same conformational cluster (3Å cutoff used for CA clustering in ROSETTA). Analysis showed that with 400 ROSETTA conformations per MODELLER model sampling is converged (data not shown). These models already show a more compact arrangement of the TM bundle transforming the outward-facing conformation into a more “channel like” structure (Figure 33). The thirty top-scoring ROSETTA models (also showing best salt-bridge conformations) were subsequently subjected to energy minimization in MacroModel as described below.

ROSETTA modeling was performed in collaboration with Dr. James Dalton. The calculations were executed by Dr. James Dalton on computers at UPF (Barcelona).

5.D.5. MacroModel minimization of ROSETTA models

A maximum of 10,000 cycles of PRCG minimization was performed on each model using the OPLS2005 force-field and a distance dielectric of 4 to approximately simulate solvent. Constraints were placed on protein dihedral angles, the position of ATP and MG atoms, distance constraints were applied to TM salt bridges, and the CL- column was frozen.

5.D.6. F508del modeling

Prime multiple-template homology modeling was used for combining the *wt* model with the F508del region extracted from the 2PZF structure of human F508del NBD1, following superimposition of 2PZF onto the model NBD1.

5.D.7. Ligand preparation for docking

Compounds were prepared for docking using the ligand preparation option in Schrödinger Virtual Screening Workflow[77]. The regularize geometry option was selected and Ionizer was used for generating tautomers and protonation states at pH=7.0±1.0.

Ligands for enrichment analysis were extracted from the Maybridge collection using the filtering option in VSW and the following criteria: MW<600 (known blockers range from 300Da to 560Da), H-Acceptors<=10, H-Donors<=5, Number of Negative groups>=1.

5.D.8. Glide docking

In glide docking, amide bond conformations were restricted to their input geometries, strain energy correction was applied and post docking minimization was performed for the top five poses. In unconstrained docking, at most five poses were reported for each ligand. During constrained docking, [N-] was added to the default charged acceptor SMARTS patterns to account for glibenclamide charge and one top scoring pose was reported for each docked compound. . An inner box size of 12×12×14 was used, centered at the mid-point between K95 and W1145 (selected following multiple tests with different box sizes and positions).

5.D.9. Molecular Dynamics Simulations

MD was performed in Desmond[145]. System setup was performed in Maestro. Of the original Cl^- column, only twenty ions were kept, filling the channel pore from top of the outer vestibule to bottom of the inner vestibule. Ion charge was reduced to -0.05 so that the total charge would not exceed that of a single passing ion. Fourteen Cl^- ions were added to neutralize the system and salt was added at a concentration of 0.15M. POPC membrane was added using the “Place on Pre-alinged Structure” option, since the model structure was based on the OPM orientation of 2ONJ. The system was solvated in SPC water. Relaxation protocol for the solvated system was prepared using the relax_membrane.py script. Simulation was carried out using the Desmond multisim utility running of 32-CPUs. Ions in the Cl^- column were harmonically constrained using a

1kcal/mol/Å force constant. Following testing, the relaxation protocol was modified by reducing the gradient convergence criterion on the second minimization stage from 5.0 to 1.0.

Trajectories were converted to PDB structures using the manipulate_trj.py script, structures were superimposed in VMD and an in-house program was used to for RMSD calculations.

6. References

(1) Woods, D. B.; Vousden, K. H. *Experimental Cell Research* **2001**, *264*, 56-66.

(2) Picksley, S. M.; Vojtesek, B.; Sparks, A.; Lane, D. P. *Oncogene* **1994**, *9*, 2523-9.

(3) Lin, J.; Chen, J.; Elenbaas, B.; Levine, A. J. *Genes and Development* **1994**, *8*, 1235-46.

(4) Bottger, V.; Bottger, A.; Howard, S. F.; Picksley, S. M.; Chene, P.; Garcia-Echeverria, C.; Hochkeppel, H. K.; Lane, D. P. *Oncogene* **1996**, *13*, 2141-7.

(5) Bottger, A.; Bottger, V.; Garcia-Echeverria, C.; Chene, P.; Hochkeppel, H. K.; Sampson, W.; Ang, K.; Howard, S. F.; Picksley, S. M.; Lane, D. P. *Journal of Molecular Biology* **1997**, *269*, 744-56.

(6) Kussie, P. H.; Gorina, S.; Marechal, V.; Elenbaas, B.; Moreau, J.; Levine, A. J.; Pavletich, N. P. *Science* **1996**, *274*, 948-953.

(7) Grasberger, B. L.; Lu, T.; Schubert, C.; Parks, D. J.; Carver, T. E.; Koblish, H. K.; Cummings, M. D.; LaFrance, L. V.; Milkiewicz, K. L.; Calvo, R. R.; Maguire, D.; Lattanze, J.; Franks, C. F.; Zhao, S.; Ramachandren, K.; Bylebyl, G. R.; Zhang, M.; Manthey, C. L.; Petrella, E. C.; Pantoliano, M. W.; Deckman, I. C.; Spurlino, J. C.; Maroney, A. C.; Tomczuk, B. E.; Molloy, C. J.; Bone, R. F. *Journal of Medicinal Chemistry* **2005**, *48*, 909-12.

(8) Lai, Z.; Auger, K. R.; Manubay, C. M.; Copeland, R. A. *Archives of Biochemistry and Biophysics* **2000**, *381*, 278-84.

(9) Zondlo, S. C.; Lee, A. E.; Zondlo, N. J. *Biochemistry (Washington)* **2006**, *45*, 11945-57.

(10) Schon, O.; Friedler, A.; Bycroft, M.; Freund, S. M.; Fersht, A. R. *Journal of Molecular Biology* **2002**, *323*, 491-501.

(11) Jabbur, J. R.; Tabor, A. D.; Cheng, X. D.; Wang, H.; Uesugi, M.; Lozano, G.; Zhang, W. *Oncogene* **2002**, *21*, 7100-7113.

(12) Abraham, M. H.; Philip P. Duce; David V. Prior; Derek G. Barratt; Jeffrey J. Morris; Taylor, P. J. *Journal of the chemical Society, perkin Transactions 2* **1989**, 1355 - 1375.

(13) Moll, U. M.; Petrenko, O. *Molecular Cancer Research* **2003**, *1*, 1001-8.

(14) Dumaz, N.; Milne, D. M.; Jardine, L. J.; Meek, D. W. *Biochemical Journal* **2001**, *359*, 459-64.

(15) Craig, A. L.; Burch, L.; Vojtesek, B.; Mikutowska, J.; Thompson, A.; Hupp, T. R. *Biochemical Journal* **1999**, *342*, 133-41.

(16) Uhrinova, S.; Uhrin, D.; Powers, H.; Watt, K.; Zheleva, D.; Fischer, P.; McInnes, C.; Barlow, P. N. *Journal of Molecular Biology* **2005**, *350*, 587-98.

(17) Showalter, S. A.; Bruschweiler-Li, L.; Johnson, E.; Zhang, F.; Bruschweiler, R. *journal of the American Chemical Society* **2008**, *130*, 6472-8.

(18) Vassilev, L. T.; Vu, B. T.; Graves, B.; Carvajal, D.; Podlaski, F.; Filipovic, Z.; Kong, N.; Kammlott, U.; Lukacs, C.; Klein, C.; Fotouhi, N.; Liu, E. A. *Science* **2004**, *303*, 844-848.

(19) Massova, I.; Kollman, P. A. *Journal of the American Chemical Society* **1999**, *121*, 8133-8143.

(20) Zhong, H.; Carlson, H. A. *Proteins* **2005**, *58*, 222-34.

(21) Kortemme, T.; Baker, D. *Proceedings of the National Academy of Sciences of the United States of America* **2002**, *99*, 14116-14121.

(22) Lee, H. J.; Srinivasan, D.; Coomber, D.; Lane, D. P.; Verma, C. S. *Cell Cycle* **2007**, *6*, 2604-2611.

(23) Dastidar, S. G.; Lane, D. P.; Verma, C. *Journal of the American Chemical Society* **2008**, *130*, 13514-13515.

(24) Engelman, A.; Mizuuchi, K.; Craigie, R. *Cell* **1991**, *67*, 1211-21.

(25) Deprez, E.; Tauc, P.; Leh, H.; Mouscadet, J. F.; Auclair, C.; Brochon, J. C. *Biochemistry* **2000**, *39*, 9275-84.

(26) Deprez, E.; Tauc, P.; Leh, H.; Mouscadet, J. F.; Auclair, C.; Hawkins, M. E.; Brochon, J. C. *Proceedings of the National Academy of Sciences of the United States of America* **2001**, *98*, 10090-5.

(27) Faure, A.; Calmels, C.; Desjobert, C.; Castroviejo, M.; Caumont-Sarcos, A.; Tarrago-Litvak, L.; Litvak, S.; Parissi, V. *Nucleic acids research* **2005**, *33*, 977-86.

(28) Guiot, E.; Carayon, K.; Delelis, O.; Simon, F.; Tauc, P.; Zubin, E.; Gottikh, M.; Mouscadet, J. F.; Brochon, J. C.; Deprez, E. *The Journal of biological chemistry* **2006**, *281*, 22707-19.

(29) Chen, A.; Weber, I. T.; Harrison, R. W.; Leis, J. *The Journal of biological chemistry* **2006**, *281*, 4173-82.

(30) Hayouka, Z.; Rosenbluh, J.; Levin, A.; Loya, S.; Lebendiker, M.; Veprintsev, D.; Kotler, M.; Hizi, A.; Loyter, A.; Friedler, A. *Proc Natl Acad Sci U S A* **2007**, *104*, 8316-21.

(31) Hayouka, Z.; Levin, A.; Maes, M.; Hadas, E.; Shalev, D. E.; Volsky, D. J.; Loyter, A.; Friedler, A. *Biochem Biophys Res Commun* **2010**, *394*, 260-5.

(32) De Rijck, J.; Vandekerckhove, L.; Gijsbers, R.; Hombrouck, A.; Hendrix, J.; Vercammen, J.; Engelborghs, Y.; Christ, F.; Debyser, Z. *Journal of virology* **2006**, *80*, 11498-509.

(33) Gupta, K.; Diamond, T.; Hwang, Y.; Bushman, F.; Van Duyne, G. D. *The Journal of biological chemistry* **2010**, *285*, 20303-15.

(34) Berthoux, L.; Sebastian, S.; Muesing, M. A.; Luban, J. *Virology* **2007**, *364*, 227-36.

(35) Dolan, J.; Chen, A.; Weber, I. T.; Harrison, R. W.; Leis, J. *Journal of Molecular Biology* **2009**, *385*, 568-79.

(36) Hare, S.; Shun, M. C.; Gupta, S. S.; Valkov, E.; Engelman, A.; Cherepanov, P. *PLoS Pathog* **2009**, *5*, e1000259.

(37) Wang, J. Y.; Ling, H.; Yang, W.; Craigie, R. *The EMBO journal* **2001**, *20*, 7333-43.

(38) Wang, L. D.; Liu, C. L.; Chen, W. Z.; Wang, C. X. *Biochem Biophys Res Commun* **2005**, *337*, 313-9.

(39) McColl, D. J.; Chen, X. *Antiviral Res* **2010**, *85*, 101-18.

(40) Cherepanov, P.; Ambrosio, A. L.; Rahman, S.; Ellenberger, T.; Engelman, A. *Proc Natl Acad Sci U S A* **2005**, *102*, 17308-13.

(41) Davis, P. B. *Am J Respir Crit Care Med* **2006**, *173*, 475-82.

(42) Du, K.; Lukacs, G. L. *Molecular biology of the cell* **2009**, *20*, 1903-15.

(43) Du, K.; Sharma, M.; Lukacs, G. L. *Nat Struct Mol Biol* **2005**, *12*, 17-25.

(44) Thibodeau, P. H.; Brautigam, C. A.; Machius, M.; Thomas, P. J. *Nat Struct Mol Biol* **2005**, *12*, 10-16.

(45) Riordan, J. R. *Annual review of biochemistry* **2008**, *77*, 701-26.

(46) Gadsby, D. C.; Nairn, A. C. *Physiol Rev* **1999**, *79*, S77-S107.

(47) Serohijos, A. W.; Hegedus, T.; Aleksandrov, A. A.; He, L.; Cui, L.; Dokholyan, N. V.; Riordan, J. R. *Proc Natl Acad Sci U S A* **2008**, *105*, 3256-61.

(48) Mornon, J. P.; Lehn, P.; Callebaut, I. *Cell Mol Life Sci* **2008**, *65*, 2594-612.

(49) Dawson, R. J.; Locher, K. P. *Nature* **2006**, *443*, 180-5.

(50) Linsdell, P. *J Biol Chem* **2005**, *280*, 8945-50.

(51) Linsdell, P. *Exp Physiol* **2006**, *91*, 123-9.

(52) Smith, S. S.; Liu, X.; Zhang, Z. R.; Sun, F.; Kriewall, T. E.; McCarty, N. A.; Dawson, D. C. *J Gen Physiol* **2001**, *118*, 407-31.

(53) Kalid, O.; Mense, M.; Fischman, S.; Shitrit, A.; Bihler, H.; Ben-Zeev, E.; Schutz, N.; Pedemonte, N.; Thomas, P. J.; Bridges, R. J.; Wetmore, D. R.; Marantz, Y.; Senderowitz, H. *Journal of computer-aided molecular design* **2010**, *24*, 971-91.

(54) Verkman, A. S.; Lukacs, G. L.; Galietta, L. J. *Current pharmaceutical design* **2006**, *12*, 2235-47.

(55) Charmm 33b1, Accelrys, Inc., San Diego, CA, 2008.

(56) Zhou, R. H.; Friesner, R. A.; Ghosh, A.; Rizzo, R. C.; Jorgensen, W. L.; Levy, R. M. *Journal of Physical Chemistry B* **2001**, *105*, 10388-10397.

(57) Michel, J.; Verdonk, M. L.; Essex, J. W. *Journal of Medicinal Chemistry* **2006**, *49*, 7427-7439.

(58) Gilson, M. K.; Zhou, H. X. *Annual Review of Biophysics and Biomolecular Structure* **2007**, *36*, 21-42.

(59) MacroModel 9.5, Schrodinger, LLC: New York, NY, 2005.

(60) Guimaraes, C. R. W.; Cardozo, M. *Journal of Chemical Information and Modeling* **2008**, *48*, 958-970.

(61) Lyne, P. D.; Lamb, M. L.; Saeh, J. C. *Journal of Medicinal Chemistry* **2006**, *49*, 4805-4808.

(62) Espinoza-Fonseca, L. M.; Trujillo-Ferrara, J. G. *Biochemical and biophysical research communications* **2006**, *343*, 110-116.

(63) Garcia-Echeverria, C.; Chene, P.; Blommers, M. J. J.; Furet, P. *Journal of Medicinal Chemistry* **2000**, *43*, 3205-3208.

(64) Fasan, R.; Dias, R. L. A.; Moehle, K.; Zerbe, O.; Obrecht, D.; Mittl, P. R. E.; Grutter, M. G.; Robinson, J. A. *chemBioChem* **2006**, *7*, 515-526.

(65) Chen, J. C.; Krucinski, J.; Miercke, L. J.; Finer-Moore, J. S.; Tang, A. H.; Leavitt, A. D.; Stroud, R. M. *Proceedings of the National Academy of Sciences of the United States of America* **2000**, *97*, 8233-8.

(66) Prime 2.2, Schrodinger, LLC, New York, NY, 2010.

(67) Schneidman-Duhovny, D.; Inbar, Y.; Nussinov, R.; Wolfson, H. J. *Nucleic acids research* **2005**, *33*, W363-7.

(68) MacroModel 9.8, Schrodinger, LLC, New York, NY, 2010.

(69) Simons, K. T.; Bonneau, R.; Ruczinski, I.; Baker, D. *Proteins* **1999**, *Suppl 3*, 171-6.

(70) Sali, A.; Blundell, T. L. *Journal of Molecular Biology* **1993**, *234*, 779-815.

(71) Shen, M. Y.; Sali, A. *Protein science : a publication of the Protein Society* **2006**, *15*, 2507-24.

(72) Bahar, I.; Atilgan, A. R.; Erman, B. *Fold Des* **1997**, *2*, 173-81.

(73) Haliloglu, T.; Bahar, I.; Erman, B. *Phys Rev Lett* **1997**, *79*, 3090-3093.

(74) Bahar, I.; Lezon, T. R.; Bakan, A.; Shrivastava, I. H. *Chemical reviews* **2010**, *110*, 1463-97.

(75) Bahar, I.; Rader, A. J. *Current opinion in structural biology* **2005**, *15*, 586-92.

(76) Christ, F.; Voet, A.; Marchand, A.; Nicolet, S.; Desimmie, B. A.; Marchand, D.; Bardiot, D.; Van der Veken, N. J.; Van Remoortel, B.; Strelkov, S. V.; De Maeyer, M.; Chaltin, P.; Debyser, Z. *Nat Chem Biol*, *6*, 442-8.

(77) Maestro 9.1, Schrodinger, LLC, New York, NY, 2010.

(78) SiteMap 2.4, Schrodinger, LLC, New York, NY, 2010.

(79) Maignan, S.; Guilloteau, J. P.; Zhou-Liu, Q.; Clement-Mella, C.; Mikol, V. *J Mol Biol* **1998**, *282*, 359-68.

(80) Molteni, V.; Greenwald, J.; Rhodes, D.; Hwang, Y.; Kwiatkowski, W.; Bushman, F. D.; Siegel, J. S.; Choe, S. *Acta Crystallogr D Biol Crystallogr* **2001**, *57*, 536-44.

(81) Halgren, T. A. *J Chem Inf Model* **2009**.

(82) 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*, 1739-49.

(83) Glide 5.6, Schrodinger, LLC, New York, NY, 2010.

(84) Phase 3.2, Schrodinger, LLC, New York, NY, 2010.

(85) Discovery Studio 2.1, Accelrys, Inc., San Diego, CA, 2008.

(86) Canvas 1.3, Schrodinger, LLC, New York, NY, 2010.

(87) Valkov, E.; Gupta, S. S.; Hare, S.; Helander, A.; Roversi, P.; McClure, M.; Cherepanov, P. *Nucleic Acids Res* **2009**, *37*, 243-55.

(88) Grobler, J. A.; Stillmock, K.; Hu, B.; Witmer, M.; Felock, P.; Espeseth, A. S.; Wolfe, A.; Egbertson, M.; Bourgeois, M.; Melamed, J.; Wai, J. S.; Young, S.; Vacca, J.; Hazuda, D. J. *Proc Natl Acad Sci U S A* **2002**, *99*, 6661-6.

(89) De Luca, L.; Gitto, R.; Christ, F.; Ferro, S.; De Grazia, S.; Morreale, F.; Debyser, Z.; Chimirri, A. *Antiviral Res* **2011**, *92*, 102-7.

(90) Smart, O. S.; Goodfellow, J. M.; Wallace, B. A. *Biophysical journal* **1993**, *65*, 2455-60.

(91) Humphrey, W.; Dalke, A.; Schulten, K. *J Mol Graph* **1996**, *14*, 33-8, 27-8.

(92) Cai, Z.; Lansdell, K. A.; Sheppard, D. N. *British journal of pharmacology* **1999**, *128*, 108-18.

(93) Alexander, C.; Ivetac, A.; Liu, X.; Norimatsu, Y.; Serrano, J. R.; Landstrom, A.; Sansom, M.; Dawson, D. C. *Biochemistry* **2009**.

(94) Gupta, J.; Evagelidis, A.; Hanrahan, J. W.; Linsdell, P. *Biochemistry* **2001**, *40*, 6620-7.

(95) Dawson, D. C.; Smith, S. S.; Mansoura, M. K. *Physiol Rev* **1999**, *79*, S47-75.

(96) Cui, G.; Zhang, Z. R.; O'Brien, A. R.; Song, B.; McCarty, N. A. *J Membr Biol* **2008**, *222*, 91-106.

(97) Hubbard, S. J.; Thornton, J. M. 1993.

(98) Kalman, M.; Ben-Tal, N. *Bioinformatics* **2010**, *26*, 1299-307.

(99) Atwell, S.; Brouillette, C. G.; Conners, K.; Emtage, S.; Gheyi, T.; Guggino, W. B.; Hendle, J.; Hunt, J. F.; Lewis, H. A.; Lu, F.; Protasevich, II; Rodgers, L. A.; Romero,

R.; Wasserman, S. R.; Weber, P. C.; Wetmore, D.; Zhang, F. F.; Zhao, X. *Protein engineering, design & selection : PEDS* **2010**, *23*, 375-84.

(100) Cui, L.; Aleksandrov, L.; Chang, X. B.; Hou, Y. X.; He, L.; Hegedus, T.; Gentzsch, M.; Aleksandrov, A.; Balch, W. E.; Riordan, J. R. *J Mol Biol* **2007**, *365*, 981-94.

(101) Sharma, M.; Benharouga, M.; Hu, W.; Lukacs, G. L. *J Biol Chem* **2001**, *276*, 8942-50.

(102) Aubin, C. N.; Linsdell, P. *J Gen Physiol* **2006**, *128*, 535-45.

(103) Schultz, B. D.; Singh, A. K.; Devor, D. C.; Bridges, R. J. *Physiological reviews* **1999**, *79*, S109-44.

(104) Wang, W.; El Hiani, Y.; Linsdell, P. *J Gen Physiol* **2011**, *138*, 165-78.

(105) Liu, X.; Dawson, D. C. *Biochemistry* **2011**, *50*, 10311-7.

(106) Maestro 8.0, Schrodinger, LLC, New York, NY, 2005.

(107) Discovery Studio 2.0, Accelrys, Inc., San Diego, CA, 2008.

(108) Sakurai, K.; Schubert, C.; Kahne, D. *Journal of the American Chemical Society* **2006**, *128*, 11000-11001.

(109) Impact 5.0, Schrodinger, LLC, New York, NY, 2005.

(110) Cheng, Y.; Prusoff, W. H. *Biochemical Pharmacology* **1973**, *22*, 3099-3108.

(111) Delano Scientific, San Carlos, CA, USA.

(112) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. *J Chem Theory Comput* **2008**, *4*, 435-447.

(113) LigPrep 2.4, Schrodinger, LLC, New York, NY, 2010.

(114) Epik 2.1, Schrodinger, LLC, New York, NY, 2010.

(115) Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. *Adv Drug Deliv Rev* **2001**, *46*, 3-26.

(116) Sperandio, O.; Reynes, C. H.; Camproux, A. C.; Villoutreix, B. O. *Drug Discov Today*, *15*, 220-9.

(117) Reynes, C.; Host, H.; Camproux, A. C.; Laconde, G.; Leroux, F.; Mazars, A.; Deprez, B.; Fahraeus, R.; Villoutreix, B. O.; Sperandio, O. *PLoS Comput Biol*, *6*, e1000695.

(118) Hare, S.; Gupta, S. S.; Valkov, E.; Engelman, A.; Cherepanov, P. *Nature* **2010**, *464*, 232-6.

(119) Rosenbluh, J.; Hayouka, Z.; Loya, S.; Levin, A.; Armon-Omer, A.; Britan, E.; Hizi, A.; Kotler, M.; Friedler, A.; Loyter, A. *J Biol Chem* **2007**, *282*, 15743-53.

(120) Krogh, A.; Larsson, B.; von Heijne, G.; Sonnhammer, E. L. *J Mol Biol* **2001**, *305*, 567-80.

(121) Tusnady, G. E.; Simon, I. *Bioinformatics* **2001**, *17*, 849-50.

(122) Juretic, D.; Zoranic, L.; Zucic, D. *J Chem Inf Comput Sci* **2002**, *42*, 620-32.

(123) Bryson, K.; McGuffin, L. J.; Marsden, R. L.; Ward, J. J.; Sodhi, J. S.; Jones, D. T. *Nucleic Acids Res* **2005**, *33*, W36-8.

(124) Petrey, D.; Xiang, Z.; Tang, C. L.; Xie, L.; Gimpelev, M.; Mitros, T.; Soto, C. S.; Goldsmith-Fischman, S.; Kernytsky, A.; Schlessinger, A.; Koh, I. Y.; Alexov, E.; Honig, B. *Proteins* **2003**, *53 Suppl 6*, 430-5.

(125) Soding, J.; Biegert, A.; Lupas, A. N. *Nucleic Acids Res* **2005**, *33*, W244-8.

(126) Jaroszewski, L.; Rychlewski, L.; Li, Z.; Li, W.; Godzik, A. *Nucleic Acids Res* **2005**, *33*, W284-8.

(127) Jordan, I. K.; Kota, K. C.; Cui, G.; Thompson, C. H.; McCarty, N. A. *Proc Natl Acad Sci U S A* **2008**, *105*, 18865-70.

(128) Hubbard, T.; Andrews, D.; Caccamo, M.; Cameron, G.; Chen, Y.; Clamp, M.; Clarke, L.; Coates, G.; Cox, T.; Cunningham, F.; Curwen, V.; Cutts, T.; Down, T.; Durbin, R.; Fernandez-Suarez, X. M.; Gilbert, J.; Hammond, M.; Herrero, J.; Hotz, H.; Howe, K.; Iyer, V.; Jekosch, K.; Kahari, A.; Kasprzyk, A.; Keefe, D.; Keenan, S.; Kokocinsci, F.; London, D.; Longden, I.; McVicker, G.; Melsopp, C.; Meidl, P.; Potter, S.; Proctor, G.; Rae, M.; Rios, D.; Schuster, M.; Searle, S.; Severin, J.; Slater, G.; Smedley, D.; Smith, J.; Spooner, W.; Stabenau, A.; Stalker, J.; Storey, R.; Trevanion, S.; Ureta-Vidal, A.; Vogel, J.; White, S.; Woodwark, C.; Birney, E. *Nucleic Acids Res* **2005**, *33*, D447-53.

(129) Katoh, K.; Kuma, K.; Toh, H.; Miyata, T. *Nucleic Acids Res* **2005**, *33*, 511-8.

(130) Guindon, S.; Gascuel, O. *Syst Biol* **2003**, *52*, 696-704.

(131) Landau, M.; Mayrose, I.; Rosenberg, Y.; Glaser, F.; Martz, E.; Pupko, T.; Ben-Tal, N. *Nucleic Acids Res* **2005**, *33*, W299-302.

(132) Stevens, T. J.; Arkin, I. T. *Protein Sci* **2001**, *10*, 2507-17.

(133) Fleishman, S. J.; Ben-Tal, N. *Curr Opin Struct Biol* **2006**, *16*, 496-504.

(134) Donnelly, D.; Overington, J. P.; Ruffle, S. V.; Nugent, J. H.; Blundell, T. L. *Protein Sci* **1993**, *2*, 55-70.

(135) Adamian, L.; Liang, J. *BMC Struct Biol* **2006**, *6*, 13.

(136) Cotten, J. F.; Welsh, M. J. *J Biol Chem* **1999**, *274*, 5429-35.

(137) Wang, Y.; Loo, T. W.; Bartlett, M. C.; Clarke, D. M. *J Biol Chem* **2007**, *282*, 33247-51.

(138) Chen, E. Y.; Bartlett, M. C.; Loo, T. W.; Clarke, D. M. *J Biol Chem* **2004**, *279*, 39620-7.

(139) Zhou, J. J.; Li, M. S.; Qi, J.; Linsdell, P. *The Journal of general physiology* **2010**, *135*, 229-45.

(140) Lewis, H. A.; Buchanan, S. G.; Burley, S. K.; Conners, K.; Dickey, M.; Dorwart, M.; Fowler, R.; Gao, X.; Guggino, W. B.; Hendrickson, W. A.; Hunt, J. F.; Kearins, M. C.; Lorimer, D.; Maloney, P. C.; Post, K. W.; Rajashankar, K. R.; Rutter, M. E.; Sauder, J. M.; Shriver, S.; Thibodeau, P. H.; Thomas, P. J.; Zhang, M.; Zhao, X.; Emtage, S. *Embo J* **2004**, *23*, 282-93.

(141) Zaitseva, J.; Jenewein, S.; Jumpertz, T.; Holland, I. B.; Schmitt, L. *The EMBO journal* **2005**, *24*, 1901-10.

(142) Lewis, H. A.; Zhao, X.; Wang, C.; Sauder, J. M.; Rooney, I.; Noland, B. W.; Lorimer, D.; Kearins, M. C.; Conners, K.; Condon, B.; Maloney, P. C.; Guggino, W. B.; Hunt, J. F.; Emtage, S. *J Biol Chem* **2005**, *280*, 1346-53.

(143) Lewis, H. A.; Wang, C.; Zhao, X.; Hamuro, Y.; Conners, K.; Kearins, M. C.; Lu, F.; Sauder, J. M.; Molnar, K. S.; Coales, S. J.; Maloney, P. C.; Guggino, W. B.; Wetmore, D. R.; Weber, P. C.; Hunt, J. F. *Journal of Molecular Biology* **2010**, *396*, 406-30.

(144) Smith, P. C.; Karpowich, N.; Millen, L.; Moody, J. E.; Rosen, J.; Thomas, P. J.; Hunt, J. F. *Mol Cell* **2002**, *10*, 139-49.

(145) Maestro-Desmond Interoperability Tools 2.4, Schrödinger, New York, NY, 2010.