Università degli Studi di Parma

DIPARTIMENTO DI FISICA E SCIENZE DELLA TERRA "MACEDONIO MELLONI"
Corso di Dottorato di Ricerca in Fisica

Structural basis for the reactivity of Nitrophorin 7 with diatomic ligands and its interaction with membranes

Coordinatore:
Chiar.mo Prof. Pierpaolo Lottici

Candidato:
Alessandro Allegri

Tutor:
Dott.ssa Stefania Abbruzzetti
Prof. F. Javier Luque (Universitat de Barcelona)

Anno Accademico 2013-2014

Ad Aurelio,
ai suoi occhi neri
ansiosi di scoprire
ciò che gli altri sanno.

A Valerio,
alla sua logica spericolata
sicura nel mostrare
ciò che lui solo sa.

The pursuit of truth and beauty is a sphere of activity in which we are permitted to remain children all our lives.
(attributed to Albert Einstein)

Abstract

The research work performed during the doctorate course included computational simulations on different protein systems.

One of these threads concerned the influence of an oxydating or reducing environment on the structural properties of human Neuroglobin and their relations with the protein's functional properties in presence and absence of a disulfide bridge. The aim of this thread, followed in collaboration with prof. Darío Estrin's group (Universidad de Buenos Aires) and with dr. Zosia Gasik, PhD student at the University of Warsaw, was to shed some light on the open question about the physiological redox state of Neuroglobin and, ultimately, on this protein's role and functions.

The present thesis reports exclusively about the main research thread concerning *Rhodnius prolixus* Nitrophorin 7, a peculiar member of a family of nitric oxide trasporting proteins. This particular protein displays a distinctive behaviour, when compared with the other, better known members of its family, in terms of NO transport and release mechanisms and of interactions with phospholipidic membranes. The computational studies, aimed at clarifying aspects such as the structure of the system of cavities and tunnels inside the protein, the pH-triggered conformational transition, the membrane interaction and its influence on the NO stocking and release, were conducted in collaboration with prof. F. Javier Luque's group (Universitat de Barcelona) and dr. Markus Knipp's group at the Max Planck Institut für chemische Energiekonversion (Mülheim a. d. Ruhr).

Contents

List of Figures

List of Tables

List of acronyms

wt wild type

APBS Adaptive Poisson-Boltzmann Solver

COM centre-of-mass

CpHMD Constant pH Molecular Dynamics

DLPC dilauroyl-PC

DLPS dilauroyl-PS

ED Essential Dynamics

FT-IR Fourier transform infrared (spectroscopy)

ILS Implicit Ligand Sampling

LFP laser flash photolysis

MD Molecular Dynamics

Mgb Myoglobin

MM Molecular Mechanics

MSMD Multiple Steering Molecular Dynamics

NO nitric oxide

NP*n* Nitrophorin *n*

NP1-4 Nitrophorins 1-4

PBC Periodic Boundary Conditions

PC L-α-phosphatidylcholine

PME Particle Mesh Ewald (algorithm)

PS L-α-phosphatidyl-L-serine

QM/MM Quantum Mechanics/Molecular Mechanics

RMSD Root-Mean-Square Deviation

RMSF Root-Mean-Square Fluctuations

Rp *Rhodnius prolixus Stål.*

RR resonance Raman (spectroscopy)

SMD Steered Molecular Dynamics

ssNMR solid state Nuclear Magnetic Resonance

Chapter 1

Rhodnius prolixus Nitrophorins

Triatomine *Rhodnius prolixus* Stål. (Rp hereafter, see Figure 1.1), also known as the "kissing bug"[1], is an insect long known (see e.g. Buxton [1]) and associated to American trypanosomiasis (Kirchhoff [2]), also known as Chagas disease, caused by parasite protozoan *Trypanosoma cruzi* hosted by the insect itself. For this reason especially, Rp and its blood-feeding process have been studied with particular interest in recent times.

1.1 A family of four

The platelet antiaggregating effect of Rp's saliva is known since the 80s[3, 4]. The first literature encounters with the specific protein(s) responsible for this effect date back to the early 90s. In Ribeiro et al. [5] the identification of a protein containing a nitrosylheme with a Fe^{III} heme reversibly binding nitric

[1]This common name is sometimes used (esp. in non-scientific or divulgative literature) to indicate Rp or other triatomines with similar blood-feeding habits; its origin probably lies in a common belief about the insect's dislike for clothed skin and its preference for biting humans on bare, uncovered areas, especially during the night, and therefore frequently on the face.

Figure 1.1: *Rhodnius prolixus* in its hemimetabolism stages from nymph to adult. Source: Wikipedia.

oxide (NO), as responsible of the release of NO at neutral pH, is made, while in Ribeiro and Walker [6] the peculiar property of releasing NO upon histamine binding, particularly advantageous for a blood-sucking insect, is investigated, to the conclusion that the protein responsible is a Fe^{III} heme protein, with a particularly high affinity for histamine binding (see also Walker [7]).

In Champagne et al. [8] for the first time there is a mention of Nitrophorins (NP1-4), as four different proteins isolated from the insect's salivary glands. Sequence and experimental similarities are also reported between NP1 and NP4, and between NP2 and NP3. An account of NP1 cloning is also given.

A previous course of research on the kissing bug had shown the presence of two peptidic anticoagulant agents, namely Prolixin-S and Prolixin-G[9]. In Ribeiro et al. [10] Prolixin-S is proved to be NP2. Sequencing and cloning of Prolixin-S/NP2 is reported in Sun et al. [11]. An account of NP2's anti-hemostatic properties is given in Zhang et al. [12]. Kinetic measurements on Prolixin-S, regarding load and release of NO, are reported in Kaneko et al. [13] and suggest that while NO loading is a pH independent process, after the protein undergoes a conformational change subsequent to Fe-NO binding, the release is, on the contrary, made faster by a higher pH.

In Sun et al. [14] the red hemoprotein labelled as RpSG I is cDNA-cloned and found to have the same N-terminal sequence as NP3, thus reconducting the research line about RpSG proteins (RpSG II being Prolixin-S, see Yuda et al. [15], i.e. NP2) to Nitrophorins.

An analysis on recombinant NP1 is provided by Andersen et al. [16], in terms of NO binding and absorption spectra, which are indistinguishable from the insect-derived protein, where the rates of NO binding and release and the Soret absorption maxima are all pH-dependent. The first data about NP1 crystals are also reported. A complete crystal structure analysis of NP1 is provided by Weichsel et al. [17], where for the first time the heme is found to be "sandwiched between strands of a lipocaline-like β-barrel, and in an arrangement unlike any other gas-transport protein discovered to date". A deeper analysis about the Fe^{III}-NO and Fe^{II}-NO forms of NP1 is reported in Ding et al. [18].

Andersen et al. [19] focus on NP4, and summarize some striking properties, such as:

- whereas NO binds almost irreversibly in globins, binding to Rp Nitrophorins is readily reversible thanks to heme stabilization in Fe^{III} form;
- whereas methemoglobin and metmyoglobin readily autoreduce to Fe^{II} forms, where NO binding is extremely tight, Nitrophorins are resistant to iron autoreduction and are thus well-suited as NO transporters;
- Nitrophorins bind NO more tightly at pH 5.0 (supposedly similar to the insect's salivary glands pH) than at pH 7.4 (the blood's pH, where the saliva is injected in the host's body), optimizing the process of NO load-and-delivery;

- Nitrophorins are able to tightly bind histamine through heme coordination once NO has been released and this appears to be the source of a strong antihistaminic activity in the insect's saliva; in addition, NP1 has been reported to have thiol oxydase activity, and NP2 an anticoagulant activity (not related to the heme moiety, but rather to the inibition of the conversion of factor X to Xa in the blood coagulation cascade);

- each of the four NPs seems to have two NO-off rates which differ among the four proteins, and this suggests that the presence of four of them is aimed at adding more functionalities and to fine-tune the release of NO over a longer period of time;

- NP1-4 share a 38% sequence identity, while NP1 and NP4 have 90% sequence identity and the next most similar couple is NP2 and NP3 with a 79% sequence identity;

- one common feature in the crystal structures of NP1 and NP4 is the presence of a buried ionizable residue (Glu55) with water molecules around it.

In Andersen et al. [20] kinetic and thermodynamic analyses for ligand binding by all four proteins are presented, together with reduction potentials. The results are discussed in the light of the crystal structures of NP1, NP2, NP4, displaying open polar distal pockets, and of NP4-NO, displaying a NO-induced conformational change that leads to expulsion of solvent molecules and complete burial of NO in a nonpolar distal pocket (see Weichsel et al. [21]). The conclusion that is drawn from these data is that the tight NO binding is due to a conformational trap rather than to particularly strong bonds (hydrogen bonds or Fe-NO).

More information about NP1-4 comes from Andersen and Montfort [22], reporting an analysis of the crystal structure of NP2 and a comparison with the other three Nitrophorins. As NP2 (and partly NP3) has a peculiar anticoagulant action, a surface region near the C-terminus and the BC- and EF-loops are held responsible for this particular feature. Another specific trait of NP2 is having higher NO association and lower NO dissociation rates than NP1,4: for this, the difference in the protein core is considered crucial: Glu53's protonation state is taken as the release switch and in the case of NP2, a deprotonated Glu55[2] would hydrogen bond with Tyr81, which lies within the core and not on the protein surface as in NP1 and NP4.

Montfort et al. [23] review several antihemostatic proteins characterized by a Lipocalin fold and found in different insects such as Rp. The paper proposes a division of eight Rp saliva proteins into four groups, of which the first includes

[2]Position 55 in NP1,4 corresponds to position 53 in NP2,3 and to position 56 in NP7. As can be seen in Figure 1.4, this is the position of a conserved Glu residue.

the four Nitrophorins[3]. Three functions are described for NP1-4:

1. NO-storage and transport; released in the target tissue, NO binds to soluble guanylate cyclase, resulting in a signal cascade that causes smooth muscle relaxation and vasodilation.

2. Histamine binding; Histamine is released by the host's mast cells as a signal for inflammation and request for immune response and wound healing.

3. Preventing blood coagulation, interfering with the cascade at the factor X maturation step. This function is reported only for NP2/Prolixin-S (see Isawa et al. [26]).

The main points this review underlines are:

- the Nitrophorin family arose through gene duplication; the NP1,4 and NP2,3 pairwise sequence similarity is followed by a functional similarity, especially with respect to the kinetics of NO binding and release;

- as Lipocalins they function binding a hydrophobic molecule at one end of the typical barrel; in NP1-4 the hydrophobic molecule is the heme, tightly packed at one end of the barrel and with ten hydrophobic contacts in addition to proximal His; this results in a distorted ("ruffled") heme, which "may contribute to the stability of the ferric oxydation state";

- NP1-4 share the same structure formed by an 8-stranded β-barrel, three short helical segments and two disulphide bonds; when NO is absent, the distal pocket is open and water-filled and loops AB and GH, at the "mouth" of the barrel (i.e. at the wider end where the heme is bound) are mobile;

- Lys88 and Lys125, which in NP1,4 are in the distal pocket and bind to the heme propionates, in NP2 are not present and this makes the distal cavity larger in the latter Nitrophorin.

In Roberts et al. [27] a refinement of the crystal structures of NP4-NO and NP4-NH_3 is offered. This allows to further appreciate the heme ruffling taking place once the cofactor is liganded. Some interesting conclusions are drawn. First of all, heme ruffling is considered instrumental to a higher stability of the ferric Fe^{III} status of the heme. Other elements that can help to achieve this function are negatively charged carboxylate groups in the proximity of the heme, or the presence of a hydrophobic cage around the NO ligand, or the

[3]The second group of proteins are the Rp aggregation inhibitors RpAI 1-3 (see also Francischetti et al. [24, 25]), and are apparently lipocalins without heme. The third group includes only one protein, an Apyrase which hydrolizes ADP to AMP. The fourth group includes only Rhodniin, an antithrombin protein.

impossibility to lengthen or dissociate the proximal histidine-heme bond, which would be required if the reduction to ferrous takes place. Another important question is raised about the role of Glu55 as a pH switch, as the presence of water molecules does not seem to change in crystals prepared at different pHs. The importance of the hydrophobic cage in terms of heme ruffling and Fe^{III}-NO stabilization has been studied in detail on NP2 and mutants by Shokhireva et al. [28] through NMR and spectroelectrochemical assays.

Maes et al. [29] report Raman spectroscopy studies on NP1-NO and NP1-CN, thereby determining bond strength and geometric parameters.

1.2 The family gets larger

Andersen et al. [30] report of the discovery of a possible novel member of the Rp Nitrophorin family, characterized by the usual Lipocalin fold, amine-binding, especially active binding serotonin and (nor)epinephrin.

Furthermore, Ribeiro et al. [31] account for a comprehensive search of the sialome (i.e. the set of mRNA plus the set of proteins found in the salivary glands) of Rp and other triatomines. While NPs seem to be peculiar to Rp, although different NPs appear at different instar stages, the NP family acquires three new members, labelled NP2A, NP3A and NP4A, according to a phylogenetic analysis of their sequences. This work also criticizes the labelling of NP5 and NP6 in Moreira et al. [32], as the sequences identified in the latter work are not found in the former and differ only by single substitutions from the sequences of NP4 and NP2 respectively. An interesting remark about polybasic sequences found in the new NPs near the C-terminal is made, as these sequences seem to be experimentally related to protein adsorption to a negatively charged membrane, such as platelet membranes due to the exposure of PS upon platelet activation, as a procoagulant surface promoting the assembly of coagulating factors.

NP7 (see Figure 1.2) first appears in Andersen et al. [33]. Recombinant NP7 is described as the only NP able to bind to anionic membranes. Upon sequence alignment, NP2-3 are the closest sequences to NP7's. The proposed function is a nonspecific blocking of membrane coagulation factor binding site, and at the same time delivering NO and binding histamine. Compared to NP1-4, more abundant, not binding to membranes and therefore remaining free in solution, NP7's action appears to be localized in the vicinity of the feeding wound, acting as platelet aggregation inhibitor, antihistamine and anticoagulant. The particular dependence of the membrane binding on salt concentration[4] and the lack of binding to PC vesicles (as opposed to PS ones) suggests that the NP7 - anionic membrane interaction is simply mediated by electrostatic contacts between positively charged residues on the protein surface and the negatively

[4]No binding is reported on 100% PC membranes, complete binding on 100% PS membranes in the absence of NaCl, and no detectable binding on 100% PS membranes at 640 mM NaCl.

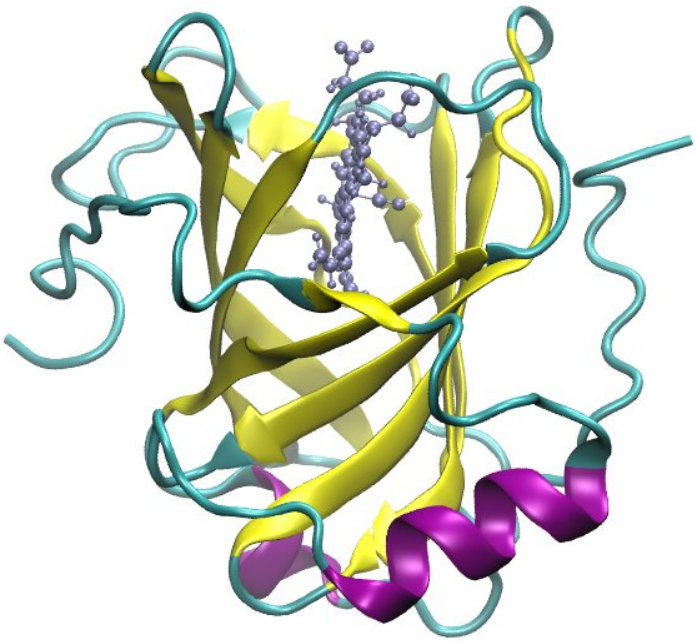


Figure 1.2: Rp NP7, cartoon backbone representation, iceblue CPK heme:CO. In yellow: the β-barrel; in purple: the α-helices; in cyan: loops. (VMD)

charged PS heads. The membrane interaction domain is pointed out as the α-helix sequentially following the Lipocalin β-barrel and before the C-terminus, where five K residues are solvent-exposed.

Meanwhile, studies on the "old" members, especially NP4, advance.

Ascenzi et al. [34] provide an interesting educational remark as NP1-4 can be used, together with other globins, namely Hemoglobins, as excellent molecular models to illustrate the completely different folds of gas storage and transport heme proteins, representing a unique evolutionary case.

Maes et al. [35], through stopped-flow spectroscopy on NP4 and AB- and GH-loop mutants conclude that the pH dependency for NO release is related to loop dynamics. An interesting remark is that the multi-phasic behaviour of NP4 NO release is not affected by mutations, and this should be detected in terms of static or dynamic active site heterogeneity (as pointed out by Nienhaus et al. [36]), but this cannot be seen in ultra-high resolution X-ray structures (see Kondrashov et al. [37]). A conformational change is reported for NP4 as follows: in the absence of NO, the distal heme pocket is open, the AB loop poorly ordered and the GH loop away from the "mouth" of the barrel. Upon NO binding, the AB and GH loops collapse into the binding pocket, ejecting solvent and packing nonpolar sidechains around the ligand. Asp30 becomes

buried and hydrogen bonds with Leu130, while Glu32, Asp35 and Asp129 and the N-terminus come together and form a network of hydrogen bonds. This, together with the fact that the NO molecule is stored in a nonpolar pocket and forms no hydrogen bonds with the protein suggests that the conformational changes are mediated through hydrophobic interactions. The distal loop is thus reported crucial for the NP4 pH sensitivity; in order to evaluate the role of the distal pocket hydrophobicity towards NO affinity, a comparison with NP2 is proposed: NP2 binds NO more tightly, displays faster association and slower dissociation than NP4, and one of the key structural differences highlighted is that whereas in NP4 Thr121, lying in the back of the distal pocket, orders a water molecule near the NO binding site, in NP2 the corresponding residue is Ile120, with the concomitant loss of the ordered water molecule.

Nienhaus et al. [36] propose IR spectroscopy to study structural heterogeneity, especially using firstly CO as a probe molecule, due to its strong IR absorbance bands in the heme-bound forms and when trapped in the protein matrix, and then physiologically relevant NO. A specific technique here replicated after studies on Myoglobins involves X-ray structures of NP4 at 100 K in presence of Xe, since hydrophobic cavities (Xe cavities) effectively trap Xe atoms in these conditions. One interesting result is that hydrophobicity and structural homogeneity of the NP4-CO distal pocket increase with decreasing pH. In terms of Xe cavities, the low temperature results show that there are no such pockets except for one in the proximal side, impossible to reach by photolyzed ligands. The spectroscopic observations here lend strong support to a ligand release mechanism that does not depend on the existence of internal cavities but is mainly based on a large scale structural rearrangement between a hydrophobic, closed, low pH structure and a hydrophilic, open, higher pH conformation. Benabbas et al. [38] confirm these findings and shed some additional light on the NP4 storage and release of NO (and CO, used in conjunction with Fe^{II}). Geminate recombination is attributed to rebinding from the distal pocket; the kinetic response of $Fe^{\mathrm{III}} - NO$ exhibits two exponential phases, the slower one interpreted as recombination from the hydrophilic open conformation, the faster one from the hydrophobic closed one. The overall off-rate of NO is seen to increase by more than one order of magnitude in the passage from pH 5.0 to pH 8.0.

More studies on NP4 include binding properties with 4-iodopyrazole and imidazole (Berry et al. [39]).

In 2005 Menyhárd and Keserü [40] report Molecular Dynamics studies (MD hereon) about the protonation states of ionizable residues in NP4, assigning to Asp30 and its protonation state a particularly relevant role for the conformational transition between the low and high pH conformers.

Maes et al. [41] analyse the properties of NP4 with non-physiological ferrous heme.

In the same year, a review of properties of the seven known NPs[5] is given in Andersen et al. [42], while Walker [7] provides a detailed analysis of the $\{FeNO\}^6$ complex in insect NPs (including *Cimex lectularius* - bedbug - NP).

In Ambrus et al. [43] the fact that NP1-4 can exist in multiple oligomerization states leads to the suggestion that their intensive, moderately pH- and ligand-dependent oligomerization has physiological implications.

An issue that Berry et al. [44] investigates is common to NP1-3 and NP7, and not to NP4 and involves the role of the N-terminus. As in *E. coli* NP4 is the only family member expressed with the initial methionine cleaved off, in account of the first residue being an alanine, the NP2(D1A) mutant is studied and found to possess some different properties from recombinant NP2-M0D1. This suggests that it is possible that the presence of M0 in all recombinant expressions of NP1-3 and NP7, which differentiates said recombinant proteins from the mature native Rp ones, are affected by the same differences. These differences are very slight in terms of reduction potentials $Fe^{\mathrm{III}}/Fe^{\mathrm{II}}$ or of the dissociation constant K_d^{III}; they are much more significant, though, in what concerns the ratio of the two heme isomers **A** and **B** (see Figure 1.3; Shokhireva et al. [45] for the identification of the two isomers in NP2, Shokhireva et al. [46, 47] in NP1 and NP4, Shokhireva et al. [48] in NP3), which have different release kinetics in different NPs, and their equilibrium relaxation times. The conclusion drawn in that paper is that "the differences in equilibrium **A**:**B** ratio and rates of NO release for each of the isomers of each of the nitrophorins provide kinetic data that are probably not representative of the nitrophorins in the salivary glands of the insect, where the proteins are likely kinetically trapped by NO binding, with non-equilibrium **A**:**B** hemin ratios. (...) mutants such as NP1(K1A) and NP3(D1A) should be prepared and investigated by the techniques used in this study. (...) Only then will it be possible to make conclusions as to the physiological consequences of kinetic trapping of non-equilibrium hemin ratio on the profiles for release of NO and binding of histamine by the physiological mixture of the four nitrophorins".

A study on the crystal structure of apo-NP4, Amoia and Montfort [49], shows that the β-barrel is not affected by the presence or absence of the cofactor, and therefore it appears to be rigid and strong enough to force distortions upon the heme cofactor itself.

Kondrashov and Montfort [50] report a comparative MD study of NP4 and Myoglobin with a free NO ligand (using a NO model after Meuwly et al. [51]). NP4 is here simulated at two different pH (5 and 7) through the different protonation states of ionizable residues: while Asp70 and Glu55 are kept protonated at both pH 5 and pH 7, Asp30 and Asp35 are considered charged at pH 7 and neutral at pH 5, and His120 and His124 are neutral at pH 7 and protonated at

[5] Actually, only five, i.e. NP1-4 and NP7, are explicitly presented in [42], probably due to a poor identification of NP5 and NP6, as seen above; this is reflected still nowadays, as in databases only those five structures or even sequences are found under the Rp NP names.

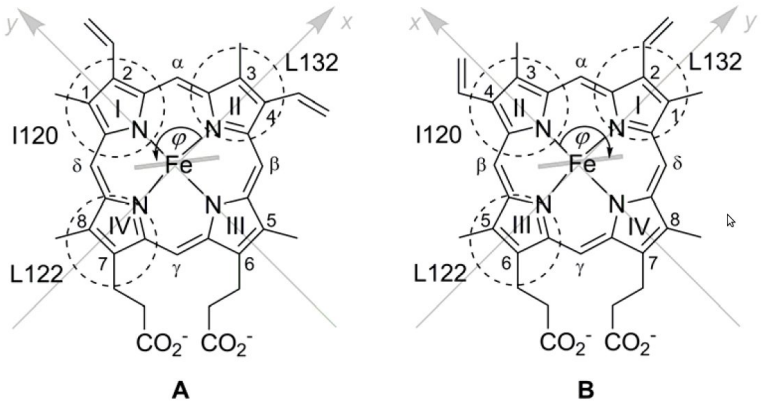


Figure 1.3: Heme in **A** and **B** orientations. Source: [55]

pH 5. This study allows to separate two different conformations for NP4 and shows NO trapping within the protein matrix, in agreement with previous kinetic data. Applying Multiple Steering MD (MSMD hereon) and Quantum Mechanics/Molecular Mechanics (QM/MM hereon) Martí et al. [52] delve deeper in the NP4 NO release mechanism and show how the Fe-NO affinity plays no role in the differential release rates, whereas the pH-triggered conformational changes affect the exit pathways of the ligand as a reversible trap.

In order to better understand the stabilization of the ferriheme, Berry et al. [53] use *wt* NP1 and several mutants of NP2 and NP4, with the substitution of ionizable residues in the heme cavity with non-ionizable ones, and then study the midpoint potentials and binding constants. The conclusion is that the stabilization of Fe^{III} in NP2 (actually NP2-D1A) "and undoubtedly (in) the other nitrophorins as well" is achieved via a large number of fairly small contributions from charged side chains. E53 and D29 seem responsible for a significant proportion of the lowering of the midpoint potential of NP2-D1A, the other residues collectively seem to give a significant contribution. L122, L132 and I120 appear as important for the ruffling of the heme, another factor increasing ferric stability, and for the **A:B** heme ratio (see also Kubo et al. [54]).

A MD and MSMD comparative study of NP2 and NP4, reported in Swails et al. [56], concludes that these two members of the NP family share a NO release control through a conformational change caused by a single conserved residue's differential protonation state at different pH, namely Asp30 in NP4, Asp29 in NP2. But they are also differentiated by the fact that the pH-induced conformational changes are more extreme in NP4 than in NP2: NP4's open, high pH conformation is far more "open" than its NP2 corresponding conformation, and this, in turn, results in the experimentally observed slower release rates for NP2 than for NP4.

	NP3	NP7	NP1	NP4
NP2	80%	60%	47%	48%
NP3		62%	43%	43%
NP7			40%	41%
NP1				90%

Table 1.1: NP1-4 and NP7 mature sequence identity matrix. (ClustalX)

	pI [a]
NP1	6.35
NP2	6.11
NP3	6.48
NP4	6.35
NP7	9.21

Table 1.2: Comparison of NP1-4 and NP7 calculated isoelectric point. [a] From Knipp et al. [58].

1.3 Nitrophorin 7

In Knipp et al. [57] NP7 and NP2[6] are compared in terms of sequence identity and sequence alignment (see Figure 1.4). In Figure 1.5 proximal His57/60 and all disulphide bridge engaged cysteines are also shown, together with the secondary structure elements. The paper reports a novel method, the previous ones being unable to yield sufficient quantities of protein for demanding laboratory use due to precipitation during refolding, for expressing NP7 in *E. coli*, purifying it and characterizing it in terms of NO, histamine and liposome binding properties. NP7 was put in contact with large unilamellar vesicles of PC and of PC:XX where XX are different phospholipids, always with a 3:1 ratio. Only when XX is a negatively charged phospholipid (PS, phosphatidic acid, phosphatidylglycerol, phosphatidylinositol, cardiolipin) does NP7 bind, while with neutral lipids no binding is reported (XX being PC, lysoPC, sphingomyelin). In contrast, none of the other forms NP1-4 does bind with any vesicle at all.

Knipp et al. [58] report the overexpression in *E. coli* of two variants of NP7: the recombinant *wild type*, with expression related M0 (MLPGEC-), and the $\Delta 1-3$ variant, always including M0 (MEC-).

In Table 1.2, the calculated pI of the NPs, as in [57], are reported: a striking difference is apparent in the isoelectric points.

An analysis about the cleavage of the signal sequence shows that for NP7

[6] As can be seen in Table 1.1, NP3 actually has the highest sequence identity with NP7, but the amount of information available on NP2 is far larger than for NP3.

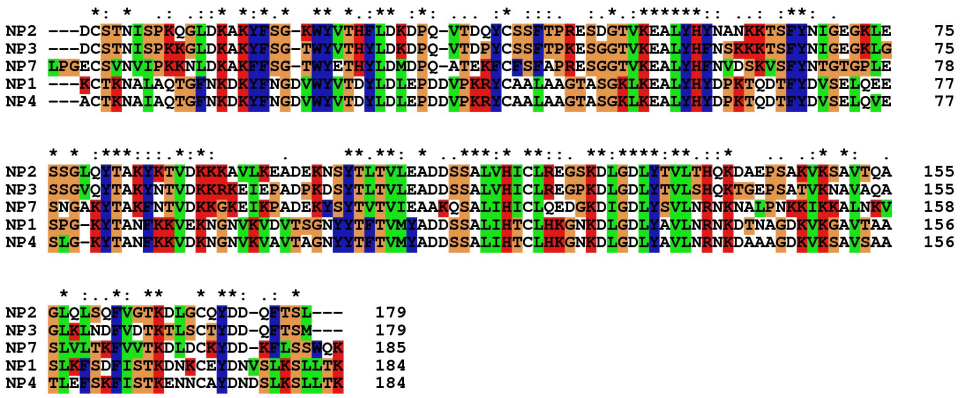


Figure 1.4: NP1-4 and NP7 sequence alignment. (ClustalX)

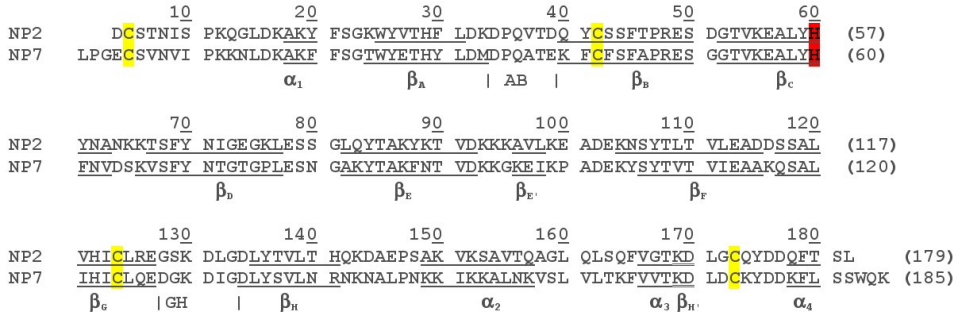


Figure 1.5: NP2 and NP7 sequences with secondary structures. Red: His57/60. Yellow: disulphide bridging Cys's. Source: Knipp et al. [57].

there is little doubt about the LPGEC- N-terminus, whereas for NP2 and NP3, up to this point considered as having a 3-residues shorter sequence than NP7, a suggestion is made that their sequences might actually include an identical VSG- head.

The conclusions of [58] characterize NP7 as follows:

1. it has the largest difference in NO association constants between high and low pH (see Table 1.3) ;
2. it has a small histamine affinity, that suggests that, unlike the other NPs, NP7 does not contribute *in vivo* to histamine sequestration (see Table 1.3);
3. it has large ^1H NMR chemical shifts for the Fe^{III} form;
4. it strongly favours the **A** orientation of the heme, unlike all other NPs;

	$\log_{10} K_{eq}^{III}$ ($\log_{10} M^{-1}$)[a]			E^o_{NO} (mV *vs.* SHE)[b]	
	NO		histamine	NO	
	pH 5.5	pH 7.5	pH 7.5	pH 5.5	pH 7.5
NP1	6.92	6.1	8.0	+154± 5	+127± 4
NP2	∼ 9.0	8.3	8.0	+49± 3	+8± 3
NP3		7.0	7.60	+73± 1	+13± 1
NP4	7.30	6.92	8.18	+94± 5	-[c]
NP7	>9	6.6	5.0	+94± 5	+109± 7
NP7(Δ1-3)	8.2	∼9.0	7.1	+157± 3	+228± 4

Table 1.3: Comparison of NP1-4 and NP7 experimental properties. [a] Equilibrium constants from Knipp et al. [58]. [b] Standard reduction potentials from Yang [59]. [c] not measurable for facile dissociation.

5. it has a unique LPG N-terminal sequence which contributes significantly to the protein fold stability and to the protein functionality.

Another interesting property exposed in He and Knipp [60] is the catalysis of the reaction

$$3NO_2^- + 2H^+ \rightarrow 2NO + NO_3^- + H_2O$$

that takes place at the ferriheme centre of NP7, suggesting that NP7 may act not only as a NO transport protein, but also as a NO producer. This reaction is dealt with in He et al. [61] through UV-visible absorption spectroscopy, electron paramagnetic resonance, resonance Raman spectroscopy and stopped-flow kinetic measurements on NP7 and NP4.

In Yang et al. [55] the peculiar role of Glu27 is investigated. Glu27, as can be seen in Figure 1.4, is only present in NP7, where in NP1-4 there is uniformly a valine. By studying NP2, NP2(V24E), NP7, NP7(E27V) and NP7(E27Q) the key role of Glu27 is identified, as its presence, in terms of steric clash since Gln27 has the same effect, is the key to the peculiar orientation of the heme (**A**) unlike all other NPs. Another feature is here reported, inferred by the different stability shown by NP7 and NP7(E27Q), that E27 must possess a $pK_a < 5$ and be therefore always charged, both at high and low pH.

In his doctoral thesis Yang [59] focuses on the characterization of NPs via NMR spectroscopy. In particular NP7 is studied in comparison with NP2. Some of the important results that are here mentioned are:

- in NP2(V24E) there are dramatic changes in the heme **A**:**B** ratio with pH, and they are explained in terms of changes in the protein structure in the vicinity of Glu53, which gets neutralized at low pH;
- the LPG sequence at the N-terminus stabilizes the fold of NP7 as absorption spectroscopy studies show that irreversible protein aggregation

starts at about 45 °C for the NP7($\Delta 1-3$) variant in front of the 52 °C of the *wt*NP7;

- dynamic light scattering experiments show that unlike NP2, which is essentially free of oligomers, NP7 contains a large fraction of oligomers;
- judging from NO binding constants (see Figure 1.3) it would seem unlikely for the NP7($\Delta 1-3$) variant to be able to release NO in blood plasma; also from reduction potentials (see again Figure 1.3) it would seem likely that the ferriheme might be reduced to ferroheme and this in turn would make NO release impossible; kinetic data (Table 1.4) show for the variant a very fast k_{on}, setting it apart from all other structures;
- therefore the presence of the LPG sequence at the N-terminus appears to have an important impact on the protein's functionality; all the same the presence of the M0 residue must be noticed in all recombinant structures (including the NP7($\Delta 1-3$) variant), and it is known that such residues in other NPs have an impact in the **A:B** heme ratio; e.g. in NP2 (see, beside the papers already cited, Berry et al. [62]) the addition of M0 causes such ratio to pass from ∼1:20 to 1:8; the importance of such residue must still be evaluated correctly for NP7;
- the NP7(E27Q) mutant precipitates at high concentrations (1mM) and could not be studied by NMR spectroscopy;
- spectra seem to indicate that the presence of Glu27 might induce some exchange process in the protein structure;
- through circular dichroism NP7(E27Q) is detected to have the same heme **A** preference as the *wt*NP7 and NP2(V24E); mixed species are detected in NP7(E27V); nonetheless, the NP7(E27Q) is much less stable than the *wt*, and this suggests the importance of the Glu27's charge for the protein fold, whereas it is the steric properties of Glu/Gln27 to be important in the selection of the heme type;
- since the NP2(V24E) mutant has a peculiar pH-dependent heme **A:B** ratio, the suggestion is proposed that the structural rearrangement, in this case, is controlled by the protonation state of Glu24; this suggestion has a particular importance for NP7, as the latter is the only family member with a charged residue in the position where all the others have a valine.

Knipp and He [63] list and describe in some detail the various functions the NPs have been found responsible for:

NO transport and release: it seems the most straightforward, but considering that the NO concentration in vertebrate blood lies in the nM range, at least one of the NPs, namely NP2 (see Table 1.3, remembering that $K_D =$

	k_{off} (s^{-1})	k_{on} ($10^6\,M^{-1}\,s^{-1}$)
NP1	2.2±0.1	1.5±0.1
NP2	0.12±0.01	33
NP3	0.08±0.01	6.7
NP4	2.6±0.1	2.3
NP7	0.61±0.13	2.4
NP7(Δ1-3)	0.50±0.02	∼500

Table 1.4: Comparison of NP1-4 and NP7 kinetic properties at pH 7.5. Source: Yang [59].

K_{eq}^{-1}, see also Table 1.4 for the tissue pH k_{on} and k_{off} values) would find releasing NO in the host's vessels quite a hard task;

histamine scavenging: this is the only other generally accepted functionality all NPs seem to have in common; yet, from experimental data (see again Table 1.3), it seems that NP7 has too low an equilibrium constant to be able to sequester histamine from the host's bloodstream; it appears then, up to this point, that the considerable translational expense for NP7 is just worth a 1:1 NO:protein release;

catalysis of NO formation from nitrite: this is apparently a functionality that NPs are able to perform *in vivo* and at neutral pH; further research is required to understand whether the ordinary blood concentration of nitrite ion is sufficient to be exploited by NPs, and how well, differentially, by the various members of the family;

interaction with high molecular weight blood components: two such components have been identified so far, corresponding to two different members of the NP family:

1. NP2 is reported to be able to directly interact with factor IXa, thus inhibiting the subsequent assembly with intrinsic factor Xase complex on the surface of anionic platelets; even if not all details of this interaction are known, this anti-blood-coagulation action was among the first to be identified in uncharacterized samples of Rp saliva;
2. NP7 is reported to be able to interact with the negatively charged (PS-bearing) external membrane of activated platelets; since the exposure of PS in activated platelets is related to the assembly of coagulation complexes and to the clearance of apoptotic cells by macrophages, this peculiarity of NP7 may allow it to target NO release on-site without risk of loss due to diffusion and blood flow,

and at the same time to reduce the negatively charged, coagulation-cascade-relevant areas on activated platelets.

In Knipp et al. [64] the two Cys-Cys disulphide bridges (C5-C124, C42-C173) are investigated as possible redox active centres.

In He et al. [65] the loss of proximal histidine coordination only detectable in Fe^{II}-NP7 and Fe^{II}-NP2(V24E) is studied as an effect of structural tension caused by the position of Glu27/Glu24 in NP7 and NP2(V24E) respectively.

Abbruzzetti et al. [66] report an extensive analysis of the Fe^{II}-CO (used as a model for the isoelectronic Fe^{III}-NO, because the slower CO binding and release rates make kinetics easier to follow) binding and release kinetics performed on NP4, NP4(D30N) and NP7 at the two *in vivo*-relevant pHs, through laser flash photolysis combined with FT-IR and RR spectroscopy, X-ray crystallography and stopped-flow kinetics.

For NP4, these studies show that there are two Xe cavities which are not populated by CO at cryogenic temperature, but become populated at room temperature, with a larger population at pH 7.5 than at pH 5.5, during LFP experiments. The kinetic scheme adopted to fit the data is reported in Figure 1.6 for both NP4 and NP7. In NP7 at least three reaction intermediates are found, corresponding to three Xe cavities. When CO is docked in these cavities, it can remain there for long times, especially at pH 5.5: this is an evidence of the particular efficiency of NP7 in storing CO (NO).

Whereas for NP7 at all pHs and for NP4 at pH 5.5 no heterogeneity was shown, the behaviour of NP4 at pH 7.5 must keep into account the coexistence of two conformers, one "open" (o) and one "closed" (c), apparent both in the high resolution crystal structures and FT-IR measurements.

Therefore, combining NP7's effective ligand in-cavity storage capabilities and its family-unique membrane binding feature, the hypothesis that one of its main functions might be to release NO at the activated platelet's surface, or near it, arises.

In [67] a four-states model to understand the pH-induced conformational transition for NP4 is proposed. Through Constant pH Molecular Dynamics, a pK_a evaluation for Asp30, the residue whose protonation state is considered the conformational switch, is obtained, with values of 8.5 and 4.3 for the closed (pH = 5.5) and open (pH = 7.5) conformations respectively. The proposed model offers an agreement with the experimentally observed pK_a of 6.5 in terms of dissociation or association and consequent conformational relaxation.

While recombinant NP7 was available since earlier, it was only in 2012 with Ogata and Knipp [68] that a crystal structure, and therefore a pdb file usable for Molecular Dynamics, was obtained. It must be said, though, that the homology model built for NP7 is in excellent agreement with this novel crystal structure.

In Knipp et al. [69] a method is described to detect the native N-terminus sequence for NP7. The method focuses in the selection and isolation of NP7

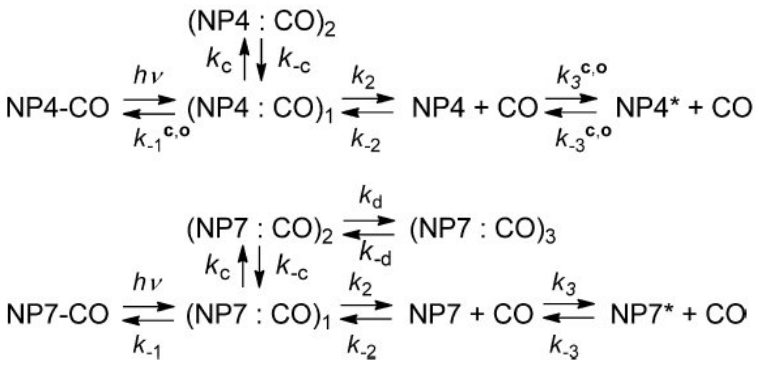


Figure 1.6: NP4 and NP7 LFP kinetic scheme. Source: Abbruzzetti et al. [66].

from Rp's saliva: the concentration of NP7 is quite low, but taking advantage of the protein's unique membrane binding feature it is possible to remove other NPs, all with a similar molecular mass, and then proceed with mass spectrometry. This method was able to confirm the difference between the recombinantly expressed NP7, with N-terminus $M_0 - LPGEC_5 -$ and the native, saliva-extracted NP7, with N-terminus $LPGEC_5 -$.

Finally, Varghese et al. [70] report a ssNMR study on NP7 precipitates, concluding that in concentrations between 0.1 mM to 2 mM oligomers are in equilibrium with monomers. The preferential oligomerization is in trimers, hexamers and higher unidentified oligomeric states. At higher concentration, no monomers are detectable. The fact that in Rp's saliva the concentration of NP7 is quite low and not yet quantified leaves the question open as whether NP7 is found there in monomeric units or at least in part oligomerized. Apparently, though, the structural properties of NP7 in oligomers are preserved. The study is an important step towards the employment of ssNMR where solution NMR or X-ray crystallography are of no use.

1.4 Open questions about NP7

This being the general picture, there are still a few points concerning NP7 that seem to require particular attention, in order to clarify aspects not well understood as yet. In particular:

from a structural point of view, how the differences with the other NPs in terms of kinetics are related to the structure and protein fold; how the length and composition of the N-terminus affects the protein structure and cavity network; how the conformational pH-controlled changes are achieved; whether the interaction with a membrane of such rigid β-barrel shaped protein affects or not the overall protein structure, and if it does,

how; what is the role for the protein fold and its conformational changes of the family-unique Glu27 residue;

from a functional point of view, what differentiates NP7 from the other NPs, what is the aim of the membrane recognition, how it is connected to the NO-release function that seems well established for all NPs.

It is to address these points that the research efforts described in the present thesis are meant, in particular using *in silico* techniques, described in Chapter 2.

In Chapter 3 the results published in Oliveira, Allegri, Bidon-Chanal, Knipp, Roitberg, Abbruzzetti, Viappiani, and Luque [71] are presented about ligand migration in NP7 and NP7(Δ1-3).

In Chapter 4 some results about the NP7(E27Q) mutant are shown, which give the opportunity to offer a consistent interpretation of the role of Glu27 for the NP7 fold.

Finally in Chapter 5 the results of MD simulations of NP7 interacting with a phospholipidic bilayer are presented which, in connection with NMR spectroscopy data, suggest that the membrane interaction might be an evolutionary strategy aimed at triggering a delayed NO release into the victim's bloodstream.

Chapter 2

Molecular Dynamics methods

2.1 Dealing with complexity - the classical way

The high number of atoms of typical protein systems[1] makes a full quantum mechanical treatment of the system impossible at present. Molecular modelling has overcome this impossibility by means of a variety of methods differing in terms of nature, depth and width of their approximation.

MD belongs to the family of classical methods. These are a set of techniques having in common the description of the atoms of biologically relevant molecules as point-masses parametrized with quantities such as fractional charge and van der Waals radius, and chemical bonds as springs connecting the atoms. The forces these atoms are subject to are therefore classical ones. This type of approximation proves accurate within the boundaries of phenomena not involving important redistributions of the electronic density of the molecules, such as the formation and breaking of chemical bonds (see [72]).

The fundamental requirements for such a classical description on a molecular basis are two [73]:

1. the representation of the energetic contributions as effective *additive* potential energies associated to simple forces

2. the transferability of such potential energies.

The first one implies that the force acting upon each atom is given by the classical equation

$$\mathbf{F}(\mathbf{r}) = -\nabla U(\mathbf{r})$$

where $U(\mathbf{r})$ is the potential energy given by a sum of contributions:

[1] NP7: 2881 atoms for the peptidic chain, 75 for the cofactor+ligand chain; within a typical 69 Å cubic solvent box including water and counterions the atom count reaches 30k. The system including a phospholipidic bilayer used in Chapter 5 is composed of more than 110k atoms.

$$U(\mathbf{r}) = \sum U_{\text{bonded}}(\mathbf{r}) + \sum U_{\text{nonbonded}}(\mathbf{r})$$

The *bonded* potential energies include all energies related to the displacement of bond distances, angles, dihedrals from their equilibrium value, whereas the *nonbonded* potential energies are usually limited to the coulombic electrostatic interactions and the van der Waals ones.

The second fundamental requirement implies that experimental data or parameters obtained through quantum-mechanical calculations can be used to develop the full expression of the model potential energy.

2.2 Force fields

The expression *force field* is used to define the set of parameters and equations needed to give a full evaluation of the potential energy of a molecular system.

There are several force fields, differing basically for the number and types of terms included and the origin of the parameters required. Naturally, the main features are the same: harmonic potentials for bond lengths and angles, Fourier series for torsions, and the nonbonded potential energy generally includes a coulombic and a Lennard-Jones contributions.

AMBER [74] and CHARMM [75] are force fields belonging to the same class: the expressions of the potential energy are very similar, the differences lying in the definition of some of the parameters. In what follows, the CHARMM force field will be used as reference mainly because it is the one that has been extensively used in the studies the next chapters are about. More in-depth information can be found in [76].

State of the art force fields are generally all-atom models: all atoms, including hydrogen atoms, are explicitly represented in the model. Some of the force fields also allow the representation of lone pairs. The shortcomings of additive force fields are mainly to be found in relation with the impossibility to fully account for electronic polarization, especially in a polar condensed phase environment such as an aqueous solution. New generation force fields are currently under development which take into account these effects.

2.2.1 Bonded potential energy terms

The bonded potential energy terms include 2-, 3- and 4-body interactions to model covalently bonded atoms.

$$U_{\text{bonded}}(\mathbf{r}) = \sum_{\text{bonds}} U_{\text{bond}} + \sum_{\text{angles}} U_{\text{angle}} + \sum_{\text{dihedrals}} U_{\text{dihedral}}$$

The two-body interaction simulates the harmonic vibrational behaviour of the bond's length:

$$U_{\text{bond}} = k\left(r_{ij} - r_0\right)^2$$

where k is the spring elastic constant (actually, its double), r_0 is the equilibrium distance and r_{ij} is the dynamical distance between atom i and atom j, covalently bonded.

The three-body interaction describes the angular harmonic vibrational motion occurring between atoms i, j and k, covalently bonded with j in the middle:

$$U_{\text{angle}} = k_\theta\left(\theta_{ijk} - \theta_0\right)^2 + k_{\text{ub}}\left(r_{ik} - r_{\text{ub}}\right)^2$$

with, in the first term, k_θ being the angular spring elastic constant (again, actually its double), θ_{ijk} the dynamical angle between vectors $\mathbf{r}_{ij} = \mathbf{r}_j - \mathbf{r}_i$ and $\mathbf{r}_{kj} = \mathbf{r}_j - \mathbf{r}_k$, and θ_0 the equilibrium angle; the second term is the Urey-Bradley term, a correction accounting for the 1-3 nonbonded cross interaction determined by nonzero values of k_{ub}.

The four-body interaction approximates torsional harmonic vibrational motions between the plane of atoms 1-2-3 (ijk) and the one of atoms 2-3-4 (jkl):

$$U_{\text{dihedral}} = k_\phi\left[1 + \cos\left(n\phi - \delta\right)\right] + k_\omega\left(\omega - \omega_0\right)^2$$

where the first term involves "proper" dihedrals, with k_ϕ as the (double of the) elastic constant, n controls the periodicity of the function, δ the phase shift; the second term accounts for the improper dihedral angles, that is out of plane bending (dynamical angle ω, equilibrium angle ω_0, double of the elastic constant k_ω), necessary to preserve planarity and chirality.

In terms of computational complexity, all the bonded terms contribute with $O(N)$ terms in the summations.

2.2.2 Nonbonded potential energy terms

The nonbonded terms include the long-range interactions between all pairs of atoms, possibly excluding those couples already involved in covalent bonds:

$$U_{\text{nonbonded}}(\mathbf{r}) = \sum_{i \neq j} U_{ij}^{\text{coul}} + \sum_{i \neq j} U_{ij}^{\text{LJ}}$$

The first term is the coulombic energy:

$$U_{ij}^{\text{coul}} = \epsilon_{14}\frac{C q_i q_j}{\epsilon_0 r_{ij}}$$

with ϵ_{14} being a scaling factor usually kept equal to 1 except when the two involved atoms are separated by three covalent bonds, in which case their dihedral torsions are already computed and the electrostatic interaction can be scaled down with $0 \leq \epsilon_{14} \leq 1$; C, ϵ_0, q_i and q_j are the Coulomb constant, the

dielectric constant and the values of the two fractional charges borne by the atoms involved.

The second term is a 12-6 Lennard-Jones potential energy accounting for the weak dipole attraction at larger distances, and strong hard core repulsion at short distances:

$$U_{ij}^{\mathrm{LJ}} = \left(-E_{\min}\right)\left[\left(\frac{R_{\min}}{r_{ij}}\right)^{12} - 2\left(\frac{R_{\min}}{r_{ij}}\right)^{6}\right]$$

where $E_{\min}$ is the depth of the energy well, $U_{ij}^{\mathrm{LJ}}\left(R_{\min}\right)$.

As far as computational complexity is concerned, long range two-point interactions contribute with $O(N^2)$ terms to the summations, therefore determining the overall computational cost. The Lennard-Jones potential, though, because of its fast fading with distance, is often truncated beyond a cutoff distance. The coulombic potential energy can be truncated as well beyond a cutoff (if both are truncated, the computational cost is reduced to $O(N)$), but as its dependency on distance is $\sim r^{-1}$, fading much more slowly than the LJ potential energy, this can lead to inaccuracies in the model. One way to overcome this is setting a large cutoff distance and using switching functions to have a smooth transition from the full-strength electrostatic potential to zero. Other ways include fast algorithms, such as the Particle Mesh Ewald algorithm, able to reduce the computational cost to $O(N \log N)$ or lower, with higher accuracy than simple truncation.

A key point for the precision of the results obtained through the application of a given force field is the accuracy of the force field parameters. Both experimental and computational techniques can be used in order to obtain said parameters. Among the former, infrared and Raman spectroscopy must be mentioned, whereas the latter consist in quantum-mechanical calculations.

An account of the parametrization for the AMBER force field can be found in Bidon-Chanal [77]. As for the CHARMM force field, the reference work is [78].

2.3 Molecular Mechanics and Molecular Dynamics

Two important techniques that make use of the force field approach are Molecular Mechanics (MM hereon), and MD. The main difference between them is the inclusion of time as a dynamical variable in the latter.

MM techniques explore the configurational space of biomolecules in search of global or local energy minima, i.e. stable configurations, regardless of time evolution.

It was in the 1950s that MD methods started to be developed. Alder and Wainwright [79] used this technique to simulate the evolution of a system of rigid spheres interacting with elastic collisions. During the following decades,

thanks to the increase in computing power and the improvements in computing techniques, MD simulations became essential tools in molecular modelling, taken by themselves or in coordination with smaller scale quantum mechanical methods[2].

MD allows to obtain structural, thermodynamical and kinetic information about biomolecules and related processes, such as conformational changes, protein folding, formation of proteic complexes, relationship between structure and functions, among others.

The starting point for MD is structural data about the molecule of interest, i.e. spatial coordinates of each atom of the molecule. Once the starting structure is known, applying Newton's laws a trajectory is obtained containing the different conformations the system occupies along the simulation time.

2.3.1 Integration algorithms

The time evolution of the system after a small finite interval Δt, once the coordinates $\mathbf{x}(\bar{t})$ are known at a given time $\bar{t}$, can be expanded in a Taylor series as

$$\mathbf{x}(\bar{t}+\Delta t) = \mathbf{x}(\bar{t}) + \dot{\mathbf{x}}(\bar{t})\,\Delta t + \frac{1}{2}\ddot{\mathbf{x}}(\bar{t})\,\Delta t^2 + o\left(\Delta t^2\right)$$

Truncating the series at the second order and disregarding the higher terms, only the positions, velocities and accelerations at time $\bar{t}$ appear in the approximating polynomial. Positions at time $\bar{t}$ are known and while accelerations are related to the dynamics of the system through Newton's second law,

$$\ddot{\mathbf{x}}(\bar{t}) = -\frac{1}{m}\nabla V\left(\mathbf{x}(\bar{t})\right)$$

in order to calculate the time evolution of the system, velocities at time $\bar{t}$ need to be evaluated as well, by a random assignment compatible with the thermodynamic conditions of the system.

Clearly, the system is an example of a many-body problem, and an approximate, numeric solution is the only possible way to deal with it.

Both the AMBER suite [80] and the NAMD program [76] use the so-called *leapfrog* velocity Verlet method for integration. The NAMD version of the *leapfrog* is described next.

Given the position $\mathbf{x}_n$, velocity $\mathbf{v}_n$ and acceleration $\mathbf{a}_n$ at timestep n, the following equations are used to obtain values for the next step:

[2]The coordination of classical MD methods and quantum mechanical models was awarded with the 2013 Nobel Prize for Chemistry to three pioneers of such multiscale models, Martin Karplus, Michael Levitt and Arieh Warshel.

$$\mathbf{v}_{n+\frac{1}{2}} = \mathbf{v}_n + \frac{\Delta t}{2}\mathbf{a}_n$$
$$\mathbf{x}_{n+1} = \mathbf{x}_n + \Delta t \cdot \mathbf{v}_{n+\frac{1}{2}}$$
$$\mathbf{a}_{n+1} = -\frac{1}{m}\nabla V\left(\mathbf{x}_{n+1}\right)$$
$$\mathbf{v}_{n+1} = \mathbf{v}_{n+\frac{1}{2}} + \frac{\Delta t}{2}\mathbf{a}_{n+1}$$

It is apparent that a very important role is played by the duration of the elementary timestep Δt. On the choice of this parameter, the precision and the computational load of the calculations depend, therefore the choice of the timestep is a compromise choice between contrasting needs.

The basic requirements is that the elementary timestep should be smaller than the characteristic times of the fastest motions the simulation is supposed to capture. Since these are the bond vibrations of light atoms, with frequencies about $10^{14}\,\mathrm{s}^{-1}$, the elementary timestep should be in the femtosecond range.

A 1 fs timestep means 10^6 iterations in order to simulate 1 ns. In order to alleviate some of the computational load, some algorithms have been developed which apply constraints to the system and allow to enlarge the timestep duration. One of these algorithms, SHAKE [81], fixes the bond lenghts involving light atoms (i.e. hydrogen atoms), allowing then to disregard their vibrational motions and to start evaluating with those of angles and torsions of light atoms, characterized by a lower frequency ($\sim 0.5\times$) and, finally, making a 2 fs timestep accurate enough. NAMD also makes use of an analytical constraint algorithm, SETTLE [82], to exclude vibrational motions of water molecules from the calculations.

2.3.2 Simulation conditions

The starting point for any simulation is always an initial configuration. This can be retrieved experimentally, through X-ray diffraction of crystals or NMR analysis of solutions, or guessed and modelled based on homology with molecules with similar composition.

Likewise, velocities must be assigned to each atom, and this is usually done randomly following a Maxwell-Boltzmann distribution so that the resulting total kinetic energy corresponds to the target thermal energy at the working temperature.

Finally, the simulation conditions must be specified somehow, the fixable variables being the number of particles (N), volume (V), temperature (T), pressure (P) and total energy (E). According to the choice of which variables are constrained, different ensembles are defined:

- microcanonical (NVE)

- isothermal-isobaric (NPT)
- canonical (NVT)

The most common ensembles in protein MD are NPT and NVT.

While keeping N and V constant is achieved in a most natural way, being these variables directly controlled in the model, the conditions on T and P require specific algorithms.

As far as temperature is concerned, one option is rescaling velocities applying the right factor to keep the overall kinetic energy at values within a certain range of the value that expresses the target temperature. Other options create a temperature coupling of the system with an external thermostat (thermal bath), acting as external thermal energy tank, subtracting or adding energy to the system as needed, by rescaling velocities at every step in function of the temperature difference between the bath and the system. The most common methods of this type are the Berendsen [83] coupling and the Nosé-Hoover [84, 85] coupling. Another option is to introduce friction with solvent molecules and random effects via Langevin dynamics.

The methods regarding pressure control are similar to those for temperature control and follow the same principles: the Berendsen barostat introduces a weak coupling with an external bath using the principle of least local perturbation. A Langevin piston [86] is the application of Langevin dynamics to pressure control, using different kinds of barostats, of which the Nosé-Hoover one is frequently used.

Simulating infinitely large systems with long range potentials is obviously impossible. The systems are necessarily confined in a finite space; the larger the more computationally demanding (the number of interacting couples is $N(N-1)$ raising the complexity to $O\left(N^2\right)$), but also the smaller the more affected by boundary effects. In order to be able to reproduce macroscopic properties in a MD simulation, the interactions with the boundaries of the confined space must be eliminated somehow, a "somehow", though, that must be aware of the complexity issues and keep the computational load under control.

A way to eliminate boundary effects is the use of periodic boundary conditions (PBC hereon). The system is confined in a barrierless box that is replicated in each direction in space. During the simulation, if an atom leaves the central box crossing one of its boundaries, the atom's image belonging to the opposite neighbouring box will automatically enter the central box, keeping constant the number of atoms. Of course the box must be large enough to avoid self-interaction of the main molecule with its own image in neighbouring boxes, especially if the main molecule is charged or has an electric dipole moment.

If PBC allow for finite, albeit "large enough", systems without boundary effects, the other side of the problem, i.e. limiting the computational complexity can be achieved only partially, as previously mentioned (see 2.2.2), through a cutoff in the evaluation of nonbonded interactions. The minimum image convention, prescribing that the longest cutoff is no longer than half the shortest

box vector, takes care of the self-interaction problem. But the electrostatic potential decays too slowly for a cutoff to be an accurate way to reduce the computational load. The most common method used to avoid this problem is the Ewald sum method [87]: the sum of all electrostatic interactions is arranged in two terms with faster convergence, the first including short range interactions (direct space) with a modified coulombic potential, and the second including the rest in Fourier space. Such method can be further improved by using Fast Fourier Transform on a discrete lattice in space (PME, [88, 89]).

2.4 Analysis techniques

2.4.1 Structural analysis

As the result of a MD run is a trajectory composed of snapshots collected regularly (e.g. every 1 ps simulated time), plus some physical quantities are calculated and saved by the simulation program with similar periodicity in auxiliary files, the amount of information that can be extracted and processed during or after a MD production run is enormous. In what follows, only the most relevant analysis performed in the context of this work shall be described.

The simplest and most immediate to analyse is information about the structure and the geometry of the molecule: RMSD, RMSF, distances, angles, etc...

Usually the calculation of the RMSD is the first step of trajectory analysis, as it provides information about structural stability, and therefore on the necessity of extending the MD run to a longer simulated time in order to achieve a more stable situation. The RMSD on a trajectory can be calculated with a variety of programs. In the present study, the tools that were used are the `ptraj` program, included in the `Ambertools` suite [80], and the `RMSDTT` extension of `VMD` [90]. It is calculated as a function of simulated time τ_j as follows:

$$\mathrm{RMSD}\left(\tau_j\right) = \sqrt{\frac{1}{N_{\mathrm{sel}}}\sum_{i=1}^{N_{\mathrm{sel}}}\left(\mathbf{x}_i\left(\tau_j\right) - \mathbf{x}_i^{\mathrm{ref}}\right)^2}$$

where N_{sel} is the number of atoms included in the selection of interest (e.g.: backbone atoms of the whole protein, heavy atoms of the cofactor, etc.), $\mathbf{x}_i\left(\tau_j\right)$ is the position vector at time τ_j of the i-th atom of the selection, and $\mathbf{x}_i^{\mathrm{ref}}$ is the position vector of the same atom in the structure used as reference, which could be a chosen frame of the same trajectory, an average structure obtained from the same trajectory or another structure (e.g. the X-ray one) provided it is alignable with the trajectory's structure. RMSD(τ_j) vs τ_j plots are usually quite informative about the quality of the simulation.

RMSF (root-mean-square fluctuations) provide a complementary point of view, as they present the same information about deviations of the positions of selected atoms from their average positions in a reference structure, this time

as a function of atom number, or residue number and averaged over simulation time.

$$\mathrm{RMSF}_i = \sqrt{\frac{1}{n_{\mathrm{frames}}} \sum_{j=1}^{n_{\mathrm{frames}}} \left(\mathbf{x}_i\left(\tau_j\right) - \mathbf{x}_i^{\mathrm{ref}}\right)^2}$$

$$\mathrm{RMSF}_k^{\mathrm{by\,res}} = \frac{1}{\sum_i w_{ik}} \sum_{i=1}^{n_k^{\mathrm{atoms}}} w_{ik}\mathrm{RMSF}_i$$

where RMSF_i is the RMSF of the i-th atom vs a reference structure, $\mathrm{RMSF}_k^{\mathrm{by\,res}}$ is the RMSF of the k-th residue as a (weighed) average of its atoms, n_{frames} is the number of frames included in the calculation, n_k^{atoms} is the number of atoms forming the k-th residue and w_{ik} is a weight that is null if the i-th atom does not belong to the k-th residue, and is equal to some non-zero value (usually the mass of the i-th atom, or simply 1) if the i-th atom does belong to the k-th residue.

RMSF plots show the different mobility of parts, secondary structures, domains of the protein and provide information useful to refine the RMSD analysis.

Other geometrical and structural quantities that can be monitored over simulated time or over frequency in a trajectory include bond distances, angles and dihedral angles: obtained usually through the application of the same programs mentioned above, they provide information about the persistence of certain structural arrangements (e.g. short distances between nonbonded atoms that might suggest the presence of hydrogen bonds or salt bridges), helping to identify, in turn, the role of some residues in the protein fold or the occurrence of conformational changes.

Another interesting type of analysis involves the "closest waters" procedure, implemented in `ptraj`. It is a way to monitor whether during the trajectory there are water molecules that remain constantly near some specified residues or not. That is, to check if there are water molecules that are stably bound to the protein, as this has an important role in neutralizing residue charge or polarity, in creating multiple fold-affecting bonds, in occluding tunnels and cavities. The procedure requires the creation of a structure with only a reduced number of water molecules, e.g. 20 out of several thousands, used as labels for the same number of water molecules that, frame by frame, happen to be nearest the specified residue(s) that need to be monitored. This way, even if the water molecules, thanks to their high mobility, are moving around the water box and the actual "identity" of those near the monitored residue changes constantly, the reduced number of labels is always assigned to those molecules in the vicinity of the sidechain of interest, and distances over time can be calculated and plotted in order to point out regularities.

2.4.2 Pocket detection

MDpocket [91] is a software that applies the fast geometrical algorithm fpocket [92] to MD trajectories in order to characterize pockets and cavities in proteins. Fpocket, in turn, makes use of Voronoi tessellations and α-spheres.

MDpocket is applied to a set of aligned snapshots from the MD trajectory; it generates a grid including all the superposed structures, and it assigns α-spheres of given maximum and minimum radii to each grid point. This produces a frequency map of the occupation of the grid point, indicating if the point is permanently accessible, transiently accessible or permanently hindered.

Some auxiliary python scripts allow the generation of coordinate files including the cavity surfaces at the desired isovalue.

The application of MDpocket on subsequent segments of the same trajectories gives information on the geometrical cavities and pockets, on their persistence. This method can be complemented with ILS (see 2.4.3) to include energetic information.

Typical settings for MDpocket are reported in Table A.3.

2.4.3 Implicit ligand sampling

Implicit Ligand Sampling (ILS) [93–95] is a technique that allows to estimate the energetics of ligand migration through the protein. By calculating the potential of mean force on a sample (small) molecule (considered in an adjustable number of possible spatial orientations) placed in every node of a suitably chosen grid, ILS produces a map that can be used to visualize energy surfaces and analysed through scripts in order to plot energetic contours of pathways chosen for the ligand.

An ILS plugin, which was used for all ILS calculations of the present studies, is included in the latest releases of VMD [90]. Its typical settings are reported in Table A.4.

2.4.4 Essential dynamics

Structural flexibility is strictly related to protein functionality (see e.g. [96–102]. Therefore the study of the structural deformation and deformability of a protein can provide relevant information about the processes involving the protein itself.

There are a few computational techniques aimed to the extraction of essential deformations from an equilibrium structure. One of them is Essential Dynamics (ED) [103]. In this case, the deformation modes are obtained by diagonalizing the positional covariance matrix. In so doing, eigenvectors are obtained reflecting the nature of the different structural deformation modes, and eigenvalues which measure the contribution of each eigenvector to the global variance.

Having different systems, or a single system in different trajectory segments, and a comparative study of deformability is required, getting the essential modes allows a direct form of comparison, through a similitude index [104, 105] obtained as a scalar product of the eigenvectors:

$$\text{overlap}(\mathbf{v},\mathbf{w}) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(\mathbf{v}_i \cdot \mathbf{w}_j\right)^2$$

where n is the minimum number of eigenvectors necessary to explain a certain value (e.g. 80% or 90% of the total variance).

Given the character of scalar product, the value of the index will be near 0 when the movements are orthogonal, and near 1 when they have a high degree of similarity.

A limit of this index lies in the supposition that in two different trajectories or in two different parts of the same trajectory the same modes would have the same weights relative to the others, which is not granted. More precise indexes have been proposed in order to overcome this and other limits [106].

2.4.5 Acidity constant determination

As in NPs pH-driven structural rearrangements happen, and these, in turn, are likely to be triggered by the protonation states of ionizable residues, methods aimed at calculating the pK_as of such residues are valuable tools. One such is the Adaptive Poisson-Boltzmann Solver (APBS) software, that can be used within the theoretical framework provided by Baker et al. [107] and the procedures described on the APBS website[3].

Starting from a MD trajectory, a procedure that can be followed includes the following steps:

1. creation of a statistically relevant collection of snapshots for each examined structures;
2. assignment (via other rough evaluation software or standard values) of a protonation state for all ionizable residues;
3. once chosen the residue whose pK_a is the object of investigation ("objective residue"), creation of a copy of each structure file of the previous point, with a modified protonation state for the objective residue;
4. conversion of all pdb files to pqr (pdb2pqr utility[4]) using suitable force field options and requesting the production of an APBS-input file;
5. insertion of cofactors, ligands or any other part of the molecule whose parameters pdb2pqr does not include, together with their charges and radii;

[3] See http://www.poissonboltzmann.org/apbs/examples/pka-calculations

[4] See previous footnote.

6. for each "ionized-objective residue" pqr file, creation of a copy of the single ionized objective residue extracted from the protein and of a copy of the whole protein with all the charges of the ionized objective residue set to zero;

7. for each "neutral-objective residue" pqr file, creation of a copy of the single neutral objective residue extracted from the protein and of a copy of the whole protein with all the charges of the neutral objective residue set to zero;

8. preparation of the input files, one per pqr file including grid dimensions and lenghts extracted from the pdb2pqr output; the key parameters are the number of grid nodes and the dielectric constant ϵ inside the protein;

9. an APBS run on every input file, each loading a different pqr file and referencing it to its "neutral objective residue" version (or to self);

10. extraction from the APBS output files of the transfer free energies; for example, should the titratable residue be Asp*n* (ordinarily charged at pH 7, while Ash*n* indicates its neutral version):

$$\Delta G_n^{tr} = \left(\Delta G_{prot}^{tr} - \Delta G_{prot-Asp}^{tr} - \Delta G_{Asp}^{tr}\right) - \\ - \left(\Delta G_{protH^+}^{tr} - \Delta G_{protH^+-Ash}^{tr} - \Delta G_{Ash}^{tr}\right)$$

where ΔG_{prot}^{tr} is the total energy of the "base" whole protein with Asp*n*, $\Delta G_{prot-Asp}^{tr}$ is the total energy of the protein with null-charged relevant residue Asp*n*, and ΔG_{Asp}^{tr} is the total energy of the single residue Asp*n*; likewise the energies in the second brackets are those including Ash*n*;

11. calculation of the shift

$$\Delta \mathrm{p}K_a = \frac{\Delta G_X^{tr}}{\ln 10 \cdot RT}$$

12. calculation of the average and standard deviation on the whole collection.

This will produce the pK_a shifts, that need to be added to the standard values in order to obtain a pK_a prediction.

In Table A.5 typical values for APBS variables, included in the input files used for the present study, are shown.

Chapter 3

Ligand migration in NP7

The present chapter presents some of the details of the research work that was presented during the 17th International Meeting “Oxygen-binding and sensing proteins” (Parma, September 2012) and published in Oliveira, Allegri, Bidon-Chanal, Knipp, Roitberg, Abbruzzetti, Viappiani, and Luque [71].

3.1 The questions addressed

The aim of the study reported in this paper is twofold:

1. to examine the influence of the three N-terminus extra residues (LPG) on the ligand association/dissociation properties;

2. to examine the structural and dynamical properties of the "closed" and "open" states representing respectively the low and high pH NP7 states, especially with reference to the ligrand migration through the protein interior.

Both questions are dealt with by combining comparative measurements of the CO rebinding kinetics to *wt* NP7 and to the Δ1-3 variant lacking the first three residues, with MD simulations of native and modified aimed ad providing structural models useful to interpret the ligand migration properties and shed some light into the structure-function relationships.

The relevant role of the LPG N-terminus sequence in NP7 has been highlighted in [58, 69]. The data shown in Table 1.3, reporting the K_{eq}^{III} for NO and the E_{NO}^{0}, and in Table 1.4, i.e. the k_{on} for NP7, NP7(Δ1-3) and NP4 (this latter does not possess the first three residues found in NP7), suffice to expose why understanding the role of the leading tripeptide is interesting.

As previously mentioned, the main difference between recombinant (*E. Coli*) NP7 and the native *wt* is the presence of Met0 in the N-terminus sequence (M_0LPGEC_5-). Due to the variants' sequence, both forms NP7(Δ1-3) (with sequence EC_5-) and Met-NP7(Δ1-3) (M_3EC_5-) could be recombinantly expressed.

The ferroheme-CO complex was chosen instead of the ferriheme-NO because (see also [66]):

1. the present study is not focused on the association/dissociation properties of the cofactor-ligand complex, but on the interactions between the released ligand and the cavity system of the protein;

2. both complexes release a diatomic gas molecule with comparable size and polarity features, therefore allowing to foresee similar interaction properties with the inner part of the protein;

3. Fe^{II}-CO has a slower binding kinetics than Fe^{III}-NO and this facilitates kinetic studies.

3.2 Methods

3.2.1 Experimental methods

Nanosecond laser flash photolysis assays were performed in the laboratories at the Università di Parma on protein samples kept at 20 °C and with a 1.0 atm or 0.1 atm CO. The absorbance spectra were measured between 10 ns and 10 ms after the laser flash. The cited paper provides all the technical details.

3.2.2 Computational methods

Initial model

For lack of experimental structures, the initial structures for NP7, *wt* and variants, were obtained by homology modeling upon the NP4 pH conformers; NP2 was also used as a model, with no significant differences. Disulfide bridges were imposed between Cys5 and Cys124 and between Cys42 and Cys173. Standard protonation states were assigned to all titratable residues except for Asp32[1], considered key in the pH conformational switch, which was therefore kept neutral at low pH and charged at neutral pH. Following the experimental results, the heme was modeled in the **A** orientation (see [58]).

While this study was conducted, diffracting crystals of NP7 were obtained [68], even though further improvement is still required to elucidate the ambiguous behaviour of the electron density, which is currently on the way.

[1] Asp32's role as the pH sensitive switch of the conformational change was suggested by analogy with well established NP4's Asp30: see [35, 40, 50, 56, 67]. Such role has been confirmed by the crystal structures at different pH displaying structures that are compatible with a hydrogen bond between Asp32 and Ile132 in the closed structures, which cannot be formed when Asp32 is charged at higher pH values. See Figure 3.1.

MD setup

Several different MD runs were performed, with different initial structures and different force fields and simulation programs. In particular, the AMBER-11 package [80] with parmm99SB [108] with specific parameters for the CO-bound heme taken from [109], and the NAMD [76] program with CHARMM22 [78] parameters.

After a 50 ns preliminary run on both open and closed structures with the AMBER-11 package, the final snapshot of each of them was used as a starting point for two AMBER and one CHARMM extended simulations and for building the models of the variants.

As the NAMD/CHARMM simulations were taken care of at the Università di Parma, the focus of the present chapter is especially on them.

The NAMD/CHARMM setup is the same as reported in A.1 (Tables A.1 and A.2).

Free ligand MD

Pockets and cavities were identified and characterized thanks to MDpocket (see 2.4.2) and ILS (see 2.4.3).

A set of MD simulations with a free NO molecule in selected initial positions and different (pseudorandom) velocities was run in order to sample the ligand's escape routes, traps, mobility inside the β-barrel. The NO parameters were adapted from [93].

For these simulations NP7 was modeled in the Fe^{II}-CO form, as the structural relaxation of the protein core subsequent to the photolyzation of the ligand happens on a longer timescale than the nanosecond scale characteristic of the ligand motions, and in this view the structure around the ligand when it is released from iron coordination should still resemble the unchanged carboxy form.

3.3 Results and discussion

3.3.1 Highlights of the experimental results

The laser flash photolysis results have been interpreted through the kinetic scheme in Figure 1.6, and the best fits to the experimental data[2] allow the determination of the kinetic rebinding constants shown in Table 3.1.

These results, extending those reported in [66], may suggest that the influence of the NP7 specific N-terminus tripeptide is negligible. This seems in contrast with some of the conclusions in [58], specifically about the drastic effect that the deletion of such tripeptide has upon k_{on}(NO), which is estimated

[2]The fits were obtained using the ODE15s function and the Minuit (CERN) package within Matlab 7.0.

	NP7		NP7(Δ1-3)		Met-NP7(Δ1-3)	
pH	7.5	5.5	7.5	5.5	7.5	5.5
k_{-1} ($10^7\,s^{-1}$)	6.9	6	3.7	6	6.9	6.1
k_2 ($10^7\,s^{-1}$)	6.5	10	5.2	13	8.0	16
k_{-2} ($10^7\,M^{-1}\,s^{-1}$)	9.7	42	6	42	5.9	33
k_3 ($10^3\,s^{-1}$)	2.6	0.9	2.6	1.7	2.7	0.97
k_{-3} ($10^3\,s^{-1}$)	3	2.3	0.5	0.5	1.6	0.1
k_c ($10^6\,s^{-1}$)	8.4	6.2	8.4	8.2	7.1	6.2
k_{-c} ($10^6\,s^{-1}$)	3	3.1	3	3.1	2.9	3.1
k_d ($10^6\,s^{-1}$)	0.88	0.85	0.88	1.6	0.96	6.5
k_{-d} ($10^3\,s^{-1}$)	7.5	9.3	7.5	4.6	4.7	8.0

Table 3.1: Rate constants based on the kinetic scheme in Figure 1.6 for CO rebinding kinetics to *wt* NP7 and variants at 20 °C. Error bars are between 10% and 20%. Source: Oliveira et al. [71].

to have a 250-fold increase. The reasons for this difference are not clear and deserve further attention, but some clues lie in the following points:

- in [58] k_{on} was determined from K_{eq} and k_{off}; the latter, in turn, was determined by the replacement of NO with excess imidazole;

- in [20] similar procedures were repeated on NP1-4, and lead to the conclusions that the constants obtained through flash photolysis and ligand displacement can differ even by three orders of magnitude, perhaps because of a structural relaxation following the replacement;

- different behaviours of the association/dissociation of CO and NO cannot be excluded in principle, and neither can a different affinity of ferri- and ferroheme for water molecules in NP7.

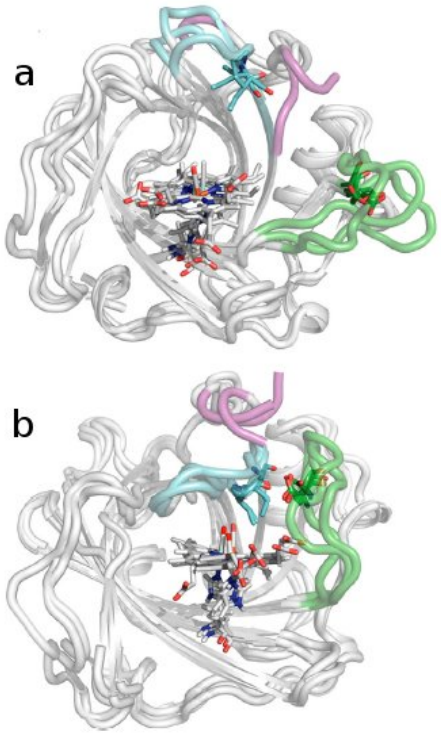


Figure 3.1: Superposition of NP7 average structures o1A, oC (a) and c1A, cC (b) (see Table 3.2, caption). Regions corresponding to the N-terminus, the AB and GH loops are shown in magenta, green and blue respectively. Source: Oliveira et al. [71].

3.3.2 Computational results

Structural analysis

An analysis of the protein backbone RMSD[3] reveals differences mainly related to the N-terminus and the AB and GH loops. The RMSD values are reported in Table 3.2.

The contribution to the RMSD differences by the N-terminus is obvious: such segment is not structured and it can interact (mainly via the positive

[3]In this case, the RMSD is calculated between the structures obtained after 1) averaging the atomic positions over the last 10 ns of the trajectories, and 2) applying energy minimization protocols to the average structures. This last step is included in order to get final structures that are dynamically plausible and that can be used as realistic representative structures of their trajectories, at least for the time window considered in the averaging process. On the other hand, the energy minimization process affects the statistical features of the average structure, in particular the one of minimizing the RMSD, in a manner that is *a priori* unpredictable. Therefore, the RMSD values calculated between energy-minimized average structures ought to be considered more as a set, for their values relative to each other, than for their absolute values.

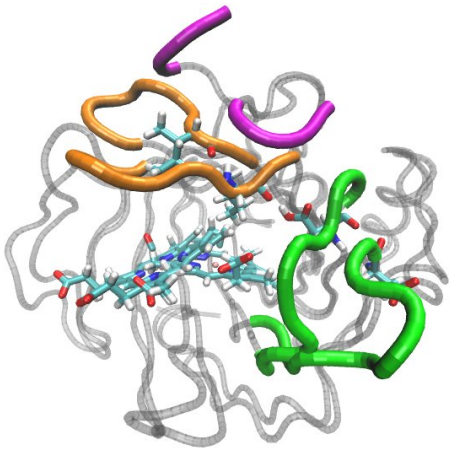


Figure 3.2: Superposition of NP7 average structures oC and cC (see Table 3.2, caption). Regions corresponding to the N-terminus, the AB and GH loops are shown in magenta, green and orange respectively. The green and orange loops are kept close to each other by the Asp32-Ile132 hydrogen bond in the closed structures, and far apart by coulombic interactions in the open structures. (VMD)

charge of the N-terminus) equally well with different residues. The role of the AB and GH loops is also clear, as they are affected by the protonation state of Asp32. It is possible to visually appreciate such differences by comparing AMBER and CHARMM average structures of the same type (Figure 3.1) and CHARMM structures of different types (Figure 3.2).

By excluding the N-terminus and the AB and GH loops from the RMSD evaluation (in italics in Table 3.2), the values decrease considerably, and group as would be expected in homogeneous open (between 1.14 Å and 1.53 Å), homogeneous closed (between 1.05 Å and 1.45 Å) and cross open-closed (between 1.43 Å and 2.11 Å).

The RMSD decrease is larger for the open structures, again as expected, since in the closed structures the AB and GH loops are kept tightly together by the hydrogen bond between Asp32 and Ile132 (Figure 3.2). The absolute values of the RMSD, especially those calculated excluding the most mobile segments, well below 2 Å, suggest that the protein core is stable and rigid, as has been experimentally shown in previous works (see e.g. [49]) for other members of the Nitrophorin family.

Again, the range of the cross open-closed RMSD values after the exclusion of the most flexible regions is anyway higher than the homogeneous values, and this justifies to trust that the different force fields have an impact that is lower than the effect associated with the open-closed conformational variability.

	o2A	oC	c1A	c2A	cC
o1A	2.40	1.76	2.63	2.72	2.88
	1.44	*1.14*	*1.98*	*1.43*	*2.11*
o2A	.	2.66	2.16	2.28	2.55
	.	*1.53*	*1.53*	*1.43*	*1.66*
oC	.	.	2.81	2.99	2.81
	.	.	*1.83*	*1.85*	*1.83*
c1A	.	.	.	1.26	1.44
	.	.	.	*1.05*	*1.19*
c2A	.	.	.	.	1.81
	.	.	.	.	*1.45*

Table 3.2: RMSD (Å) between the backbones of the *wt* energy-minimized structures obtained averaging the last 10 ns of MD trajectories. In italics: RMSD obtained excluding the N-terminus tripeptide and the AB and GH loops. "o1A" stands for "open, 1st simulation, AMBER", while "cC" for "closed, CHARMM". Source: Oliveira et al. [71].

3.3.3 Cavities

In [66], as previously reported (1.3), different kinetic schemes were proposed for NP4 and NP7 (Figure 1.6). In particular, a more complex scheme for NP7 is to be adopted in order to explain a longer residence of the CO ligand in the protein interior following photodissociation. In order to better understand the migration paths followed by the ligand, the identification of the cavity system within the protein was pursued with MDpocket (see 2.4.2) on subsequent 20 ns-long windows. The reason to choose separate, non-overlapping time windows (belonging to the same stability regions of the MD trajectories) is the possibility, which is given in so doing, to verify the reproducibility of the cavities.

Three cavities, shown in Figure 3.3, are found both in the open and in the closed structures and in simulations run with both the AMBER and the CHARMM force fields through the MDpocket isocontour frequency maps, allowing to identify how often during a MD simulation segment the cavities are connected into a single tunnel.

The first cavity (denoted site 1) includes two separate docking sites (1a and 1b), defined by the "back" side of the heme[4] and by residues Phe43, Leu58, Tyr107, Ile123, and Leu125.

The second cavity (site 2), further down the barrel, is delimited by Phe20, Phe45, Pro47, Glu56, Tyr84, Val109, Ile121 and Leu139.

[4]In Figure 1.3/**A**, the "back" side is identified by the α bridge between pyrrols I and II. The γ side is the "front" side, as the propionates stemming from pyrrols III and IV are protruding from the "mouth" of the protein's barrel.

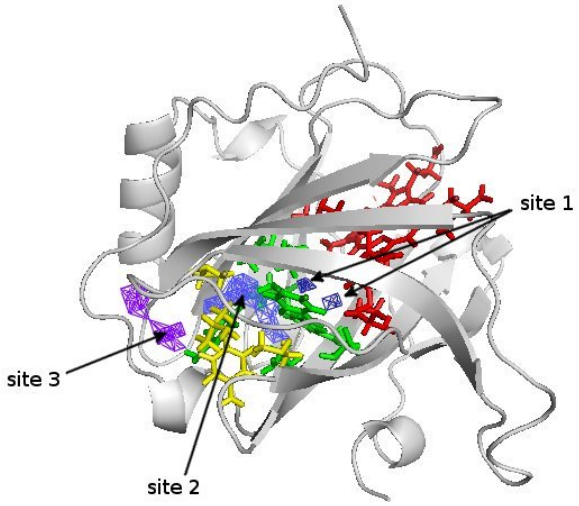


Figure 3.3: Identification of sites 1, 2 and 3 in NP7 (*wt* closed structure, isovalue: 0.75 frequency). (MDpocket/Pymol)

Site 3, further again, is lined by Phe20, Phe21, Val54, Leu77 and Val111.

Most of the cavity-defining residues are hydrophobic (Ile123 and Leu125, in site 1, also define the ligand cage in the heme distal cavity), with the noteworthy exceptions of polar residues, and even charged Glu56, all engaged in interactions that are crucial for the overall protein fold, and thus neutralizing their polar character (see [110]).

Implicit Ligand Sampling (see 2.4.3) was applied on the same 20 ns-long windows along the trajectories. In this case, the ILS isocontour maps are isoenergetic surfaces.

For the closed *wt* protein, a continuous linkage between sites 1 and 3 is visible at an energy isocontour of 3.0 kcal mol^{-1} (see Figure 3.4), which suggests that a gaseous diatomic ligand should easily exchange between the two sites. At the same energy value there is no continuous linkage between the same sites in the *wt* open protein. Such linkage becomes apparent only at a much higher energy value, about 15 kcal mol^{-1}, which would indicate that a free gaseous molecule should be mainly confined in site 1 or released to the solvent at high pH.

These results seem to point out that while in the open form the three sites only occasionally are connected into a single tunnel running inside the protein and connecting the heme environment with the "back" end of the barrel, the presence of such single tunnel, although not constant, is significantly more frequent in the closed form.

The overall structure of NP7(Δ1–3) shows a large resemblance with the

tunnel	cC	cC-Δ^a	oC	oC-Δ^a
upper tunnel[b]	65%	40%	55%	40 %
back tunnel[c]	70%	60%	50%	45 %
direct exposure[d]	60%	90%	70%	75 %

Table 3.3: Tunnels in *wt*NP7 and variants. [a] cC-Δ and oC-Δ are the $\Delta 1-3$ variants obtained with the CHARMM force field from the closed and open *wt* structures respectively. [b] The upper tunnel connects the distal pocket with the exterior through strands G and H. [c] The back tunnel connects the distal cavity with the "back" of the protein along the barrel axis, joining together sites 1, 2, 3. [d] Direct exposure of the ligand to the solvent through the "mouth" of the barrel. (MDpocket)

	o2A	oC	c1A	c2A	cC
o1A	0.77	0.81	0.53	0.53	0.55
o2A	.	0.71	0.51	0.49	0.49
oC	.	.	0.52	0.53	0.59
c1A	.	.	.	0.90	0.76
c2A	.	.	.	.	0.81

Table 3.4: Similarity indexes between the inner cavities in *wt* trajectories according to Tanimoto's distance. Source: Oliveira et al. [71].

wt protein, especially in the AMBER simulations, with RMSD values for the backbone atoms of 1.52 Å and 1.42 Å for the open and closed species, respectively. For the CHARMM simulations, the values are higher between the open structures (3.78 Å) and lower for the closed ones (1.19 Å).

The MDpocket analysis also supports the topological resemblance between the inner cavities in both *wt* NP7 and its variant. Thus, the Tanimoto's distances (see [111]) between *wt* and NP7(Δ1-3) were 0.65 (open) and 0.74 (closed), whereas the cross (open versus closed) similarity index was 0.54, thus mimicking the trends found for *wt* NP7 in Table 3.4.

3.3.4 Free ligand diffusion

In order to clarify what the influence of the partially pervious tunnel upon the ligand migration might be, a set of free ligand MD simulations was run.

In the complete protein-heme-CO structure[5] (starting from the last snap-

[5]The reason for choosing the same NP7:CO structure adding a free NO molecule lies in the fact that it is most probable that, because of the characteristic times of flash-photolysis, diatomic ligand motion and protein structural relaxation, a flash-photodissociated NO molecule finds itself

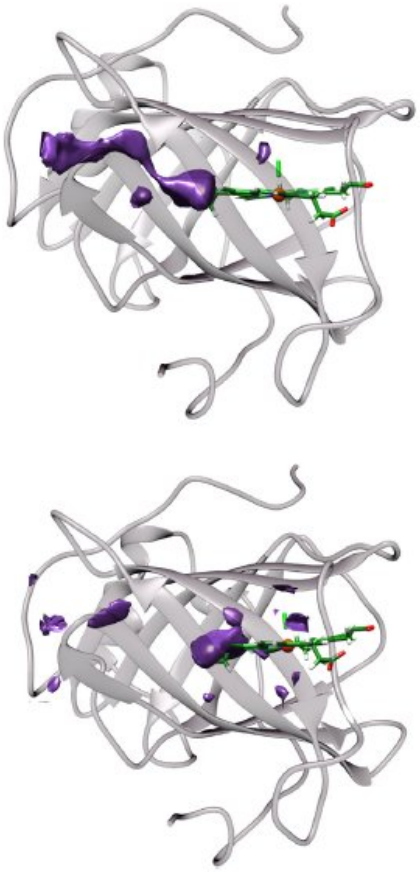


Figure 3.4: ILS energetic profiles linking inner cavities in the closed (top) and open (bottom) forms of *wt*NP7. The isocontour corresponds to an energy value of 3.0 kcal mol^{-1}. Source: Oliveira et al. [71].

shot of the trajectories for the open and closed, *wt* and Δ1-3 variants), a single NO molecule was inserted in four different positions (named A, B, C and D; see Table 3.5, Figure 3.5), and for each of these four situations, five replicas, differing in the random seed for initial velocity generation, were produced. Each simulation covered a 20 ns time span, following the ordinary energy minimization and equilibration steps.

As far as closed *wt* NP7 is concerned, the positions sampled by NO in the interior of the protein reflect the topological features of the cavities identified from MDpocket analysis. NO molecules primarily populate sites 1a, 1b and 3, whereas a lower occupancy is found for site 2, suggesting a lower effective

in a still unrelaxed protein, i.e. as if the protein conformation was frozen at the heme-ligand bond breaking instant.

ID	position	identifying residues
A	end of back tunnel (site 3)	81 15 17
B	beginning of back tunnel (site 1)	heme 121 123 135 139
C	beginning of side tunnel	30 32 37 132 135 heme
D	N-term	1 34 36

Table 3.5: Initial positions for free ligand simulations.

	pos. A		pos. B		pos. C			pos. D		
	in[2]	ex[3]	in[2]	ex[3]	ee[1]	in[2]	ex[3]	ee[1]	in[2]	ex[3]
c*wt*C	0.8	0.2[4]	0.8	0.2[4]	1.0			1.0		
o*wt*C	0.2	0.8[4]	0.2	0.8[6]			1.0[6]		0.2	0.8[5] (1:3)
c(Δ1-3)C	0.8	0.2[4]	1.0			0.2	0.8[7]	1.0		
o(Δ1-3)C	0.4	0.6[5] (2:1)	0.8	0.2[6]		0.2	0.8[6]	1.0		

Table 3.6: Free ligand exit/residence after 20 ns. [1] Exit during equilibration (no MD run performed). [2] Inside the protein after the 20 ns run. [3] Exit during MD run. [4] Exit only through the "back". [5] Exit partly through the "back" and partly through the "mouth". [6] Exit only through the "mouth". [7] Exit partly throuth the "mouth" and partly through a side tunnel (1:3).

volume for accommodating the ligand in this latter cavity. The trajectories that started from position D led to ligand egression in all cases (see Table 3.6). For position C, the ligand was either released to the solvent during equilibration or migrated through the docking sites, and in few cases it was released to the solvent from site 3 through the back of the protein (near residues Leu15 and Gly81). Finally, the trajectories started from positions A and B led to either retention of NO in the interior of the protein (70% of cases) or to egression through the back of the protein (30% of cases).

The distribution of NO between sites 1 and 3 is well reflected in Figures 3.6, 3.7, obtained by plotting the distribution of distances of the position occupied by the free ligand molecule (more precisely, by the N atom) from a reference atom along the trajectory time (i.e. Gly81's C_α[6]). Speaking of the whole set

[6]The most common choice for such internal distance parametrization is the heme iron atom. In this case it was not the most suitable one, as the heme undergoes displacements during the trajectory, while the most reliably rigid part of the system is the β-barrel. Gly81 is obviously a no-sidechain residue, its position lies on a tight loop at the very "back" end of the barrel, more external than site 3, and for these reasons its C_α atom was deemed the best choice as a reference point to parametrize NO's position in the protein's interior. In addition, this choice is supported by the structural similarity of the barrel, which allows to define a common reference frame.

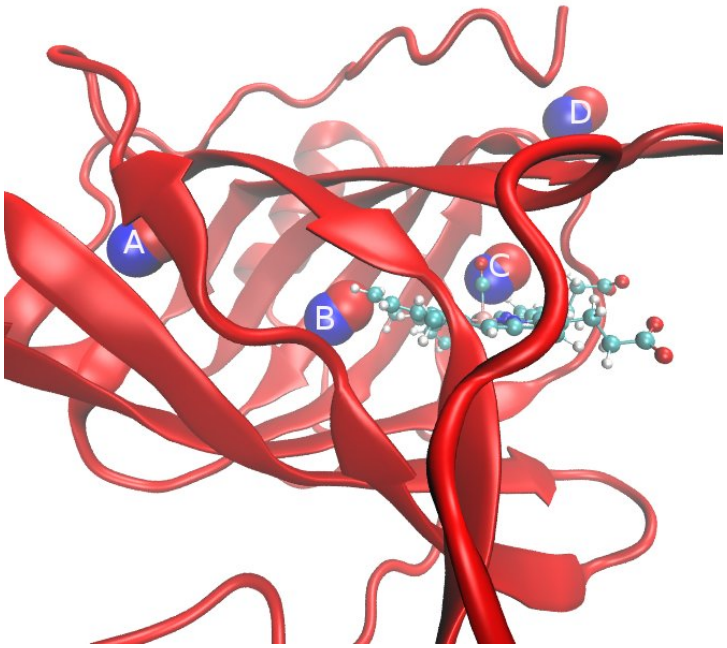


Figure 3.5: Starting positions of free ligand simulations. The labels refer to those given in Table 3.5. (VMD)

of simulations (i.e. including both the AMBER and the CHARMM results) the trajectories started from position A (site 3) sample regions to position B (site 1) and vice-versa. The migration of the ligand between sites 1 and 3 suggests the existence of small energy barriers for diffusion through the interior of the β-barrel, in agreement with the linkage between sites 1 and 3 found from ILS calculations. Finally, the lack of the peak corresponding to site 2 suggests that it is only transiently populated in the exchange pathway between sites 1 and 3. Accordingly, site 2 should not be treated as a stable docking site, but rather as a temporary site that bridges cavities 1 and 3.

For the open *wt* structures, all the trajectories started from positions B, C and D released NO to the solvent through the front side (distal cavity) of the protein, in agreement with the physiological function of releasing NO at the plasma pH. Only very rarely did NO migrate through the cavities and exited through the back of the protein (site 3). When the simulations started from po-

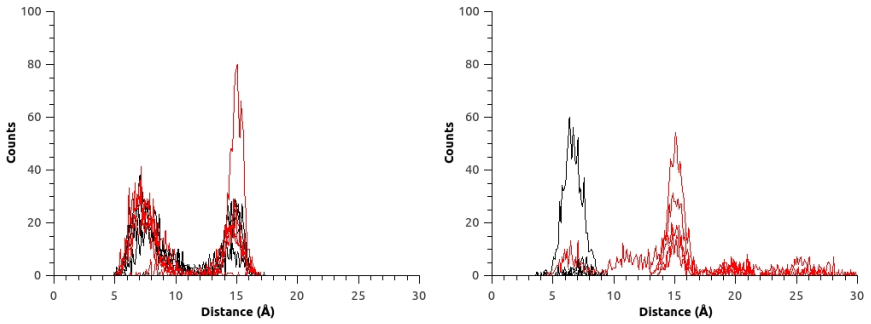


Figure 3.6: Free ligand distance (between NO's N atom and Gly81's C_α) distribution for *wt* NP7. Closed structure on the left, open on the right. Starting position A/3: black; starting position B/1: red. Source: Oliveira et al. [71].

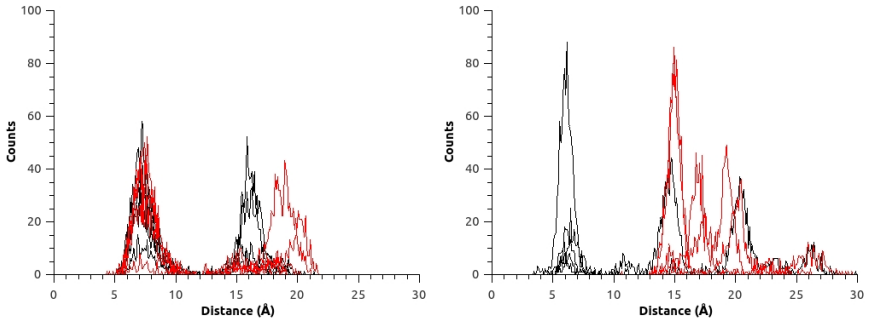


Figure 3.7: Free ligand distance (between NO's N atom and Gly81's C_α) distribution for NP7(Δ1-3). Closed structure on the left, open on the right. Starting position A/3: black; starting position B/1: red. Source: Oliveira et al. [71].

sition A (site 3), however, the ligand was released to the solvent only through the back of the protein. In contrast to the closed structures, the ligand is mainly localized in site 3, indicating the existence of enhanced barriers for ligand diffusion to site 1, in agreement with the larger isoenergy contour required to link sites 1 and 3 from ILS calculations. This is supported by the larger thermal fluctuations of the residues that define the distinct cavities in the β-barrel, as can be noticed in Figure 3.8, which compares the RMSF in both open and closed states. Thus, the lower flexibility of the β-barrel residues in the open state likely contributes to the larger barrier for ligand exchange between docking sites.

Free ligand MD simulations run on NP7(Δ1-3) using the same set of starting positions and simulation protocol confirmed the results reported for *wt*

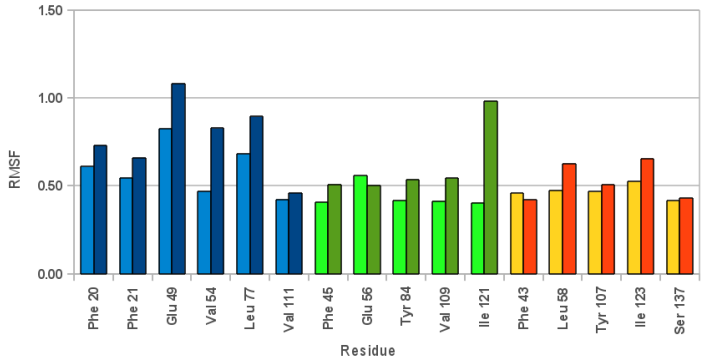


Figure 3.8: RMSF (Å) for the residues defining the internal cavities of *wt* NP7. Cavities 1, 2 and 3 are respectively blue, green and yellow colour-coded. For each residue, the dark and bright columns refer to the closed and open structures respectively.

NP7. In particular, the NO distribution profiles confirmed the enhanced diffusion of the ligand between sites 1 and 3 for the closed NP7(Δ1–3) compared to the open species, where ligand migration between inner cavities seems to be hindered.

Overall, these results allow us to suggest that deletion of the first three residues at the N-terminus does not introduce drastic structural alterations in NP7's cavity system. This supports the existence of a common pattern for ligand migration in both *wt* NP7 and NP7(Δ1–3), in agreement with the similarity in the kinetic rate constants reported in Table 3.1.

3.3.5 Comparison with NP4

In [66] two different kinetic schemes for ligand migration in NP4 and NP7 (see Figure 1.6), are proposed, based on new experimental data and on previous works (e.g. [52, 56]). In order to verify the extent of the differences in terms of cavities between the two members of the NP family, the same type of MDpocket and ILS analysis was performed on NP4 (a more precise account of which can be found in the [71] paper).

The most relevant difference between the two proteins is that NP7's site 2 is absent in NP4. This may be interpreted as the combined effect of the variants Phe20(NP7) ⟷ Tyr17(NP4), Val109(NP7) ⟷ Phe107(NP4) and Gly75(NP7) ⟷ Leu74(NP4). While the two former variations reduce the free space in the middle of the β-barrel, the latter displaces the position of Glu55 in NP4 (Glu56 in NP7) towards the interior and modifies the positions of Tyr82 and Tyr105 in NP4 (Tyr84 and Tyr107 in NP7), facilitating the opening of a small cavity between these residues in NP4 (but absent in NP7).

The lack of site 2 in NP4 should hinder the ligand exchange between the equivalent of sites 1 and 3 in NP4 compared to NP7. This agrees with the findings from previous MD simulations of NP4 where NO did not migrate to the interior of the protein, see [50, 52], even though simulations were extended for more than 100 ns. Furthermore, spectroscopy at cryogenic temperatures has shown that NO and CO rebinding to the heme occurs exclusively from the distal pocket [36]. Finally, the existence of the two docking sites in the vicinity of the heme (precisely next the ligand binding site on the distal side of the heme, and under the heme, in the proximal pocket) may explain the existence of the two phases seen in the kinetic trace of NO rebinding kinetics to ferrous NP4, as reported by [38]. While the first phase can be assigned to the rebinding directly from the distal pocket site, the second phase (geminate amplitude <5%), which is insensitive to the variation of the pH, may be ascribed to migration of photolyzed ligands to the proximal site.

3.3.6 Functional implications

As shown above, experimental results ([36, 38, 50, 52, 66]) benchmark the picture provided by the MD simulations: there is a hypothetical kinetic mechanism with an extra docking site in NP7; this mechanism can be put in correspondence with the differential availability of docking site 3, reachable from site 1 via transient site 2 in NP7 but not in NP4.

This suggests that NP7 might be functionally linked to the existence of two pathways for ligand migration. One corresponds to the release through the distal cavity, which is greatly enhanced upon transition from closed to open states due to the breaking of the hydrogen bond between Asp32 and Ile132 (see Note 1 and Figure 3.1), and the concomitant opening of loops AB and GH. A second pathway would involve the access through site 3 at the back of the protein, near residues Leu15 and Gly81.

The question has been often raised why Rp provides a bundle of NPs in its saliva instead of providing just one NP. The various isoproteins seem to address different targets in the host tissue, but in addition it seems that the proteins have different mechanisms to provide their load to the specific target. Keeping in mind that most of the NP7-specific, positively charged residues are located on the “back” of NP7, and that this protein exhibits the unusual property of targeting negatively charged phospholipid membranes [33, 57, 58], it can be speculated that the interaction with the membrane may alter the local structure at the back of the protein and thus facilitate the exchange of ligands between membrane and protein through internal diffusion from the heme to site 3.

With regard to the functional role of the NP7-specific N-terminus, the results do not allow to provide a consistent conclusion. The computational structural data and also the kinetic data derived from CO rebinding studies do not disclose fundamental distinct trends between NP7 and NP7(Δ1–3). However, the NO distribution profiles show significant differences and, thus, suggest a

non negligible influence of the Leu-Pro-Gly tail, as noted in the significant differences found in thermodynamic parameters K_{eq} and E^o obtained for both forms (see Table 1.3).

3.4 Conclusions

Summarizing the main findings already mentioned above, the conclusions that can be drawn from this section of the research work can be listed as follows:

- the differences caused by the choice of one force field instead of another are small, probably within the conformational variability allowed for the protein dynamics by both force fields;
- the existence of three main cavities in the interior of the β-barrel is suggested, which, as expected, are lined with hydrophobic residues; the existence of site 2 would make the passage between site 1 and site 3 possible for a small ligand;
- this axial tunnel, even if not permanently defined, is in contrast with the topology of inner cavities in NP4 and provides a structural basis to the enhanced internal gas hosting capacity in NP7 and to the absence of ligand rebinding from secondary sites in NP4;
- site 3 is particularly interesting as it lies at the "back" of the protein, right where an accumulation of positively charged Lys residues would suggest the interaction with negatively charged membranes can occur, and this, in turn, leads to the speculation of an interconnection between the processes of docking to cell surfaces and NO release, probably via site 3.

As NO delivery through protein-membrane interaction is an interesting hypothetical mechanism for NO signaling, future studies will be conducted in order to verify and define its actual occurrence. The availability of experimental structural information of NP7 might shed light into the functional implications of the N-terminus.

Chapter 4

Glu27 and the conformational transition between open and closed states

4.1 The questions addressed

In [67] (see section 1.3) through CpHMD a pK_a evaluation for Asp30 in NP4 was obtained, ranging between 8.5 and 4.3 for the closed (pH = 5.5) and open (pH = 7.5) conformations respectively. The proposed model offers an agreement with the experimentally observed pK_a of 6.5 in terms of the degree of ionization of Asp30 and the consequent conformational transition.

In [55] (see section 1.3) the role of E27, unique occurrence in NP7 throughout the family where all the other members have a valine (Figure 1.5), was analysed by mutating this residue to different choices. E27's role can be set in relation with the heme's position and orientation. Since the heme's position is one feature that seems to distinguish the open and closed conformations, E27 can ultimately influence the ionization of Asp32, and hence the conformational transition.

In order to clarify what influence E27 produces upon the protein structure, an experimental and computational study was performed in order to compare *wt*NP7 structures and E27 mutants, namely NP7(E27Q), NP7(E27V) and purely computational NP7(E27prot) i.e. with E27 modelled as protonated.

4.2 Methods

4.2.1 Experimental methods

Nanosecond LFP assays were performed at Università di Parma on *wt*NP7 and NP7(E27Q) at 20 °C temperature and 1 atm and 0.1 atm CO pressure and at both pH 5.5 and pH 7.5. The experimental assays were repeated extending the

structure	action
E27p1	patch GLUP 27[1]
E27p2	E27:OE1↔E27:OE2[2], patch GLUP 27[1]
E27Q1	GLU→GLN[3], Q27:OE2→Q27:NE2[3]
E27Q2	GLU→GLN[3], Q27:OE1↔Q27:OE2[2], Q27:OE2→Q27:NE2[3]
E27V	E27's sidechain deleted[4], GLU→VAL[3]

Table 4.1: E27p's and E27Q's preparation, following hydrogen atoms' deletion and preceding the application of VMD's autoPSF extension to recreate the correct bonds and place new hydrogen atoms in suitable positions. [1] Forces the structure-building autoPSF plugin to add a hydrogen atom to otherwise normally ionized Glu27:OE2 atom. [2] Manual swap between two atom names on the pdb file. [3] Manual renaming of all atoms of a residue or of a single atom, in the pdb file. [4] Manual deletion of the sidechain in the pdb file, from C_β onwards.

timescale to the ps range by using an ultrafast ps pump and probe set available at the Politecnico di Milano (Prof. Giulio Cerullo) in order to understand the geminate rebinding phase and evaluate the rebinding kinetics as a whole, and therefore also estimate in a more precise way the bimolecular fraction in different conditions.

4.2.2 Computational methods

Building the structures, getting the trajectories

Starting from a snapshot in a stable segment of NP7c's trajectory, (see Chapter 3) two versions of NP7(E27p) (named E27p1 and E27p2), two versions of NP7(E27Q) (named E27Q1 and E27Q2) and one version of NP7(E27V) (named E27V) were prepared. The reason for the double structure for all variants except the E27V is the possibility to protonate/mutate either E27:OE1 or E27:OE2. Therefore the procedure was the following:

i removal of all hydrogen atoms, water molecules, ions in the pdb file of the chosen frame;

ii actions described in Table 4.1;

iii rebuilding of structure file: application of patches, insertion of the system in a water box and addition of counterions (A.1.1).

The usual 3-step energy minimization process then took place as in A.1.2, and after that the 2-step equilibration procedure A.1.3.

Each production dynamics run, set as in A.1.4, covered 150 ns with 2 fs timesteps. They were performed by NAMD2.9 on 128 processors in 8 ns segments.

Looking for the acidity constants

The procedure described in 2.4.5 was followed in order to verify the pK_a for Asp32.

1. the collection of snapshots was chosen as 100 frames sampling 10 ns-long intervals along each trajectory;

2. the standard protonation state at pH 7.0 was assigned to all ionizable residues and termini;

3. the “objective residue” was Asp32 (Ash32);

4. the CHARMM force field was chosen;

5. insertion of the heme cofactor in all pqr files, coordinates copied in from the original pdbs, plus the CHARMM parameters (charge, radius), as the pdb2pqr does not include the heme’s parametrization among its features;

6. for each "Ash32" pqr file, creation of a copy of the single Ash32 residue extracted from the protein and of a copy of the whole protein with all the charges of the Ash32 residue set to zero;

7. for each "Asp32" pqr file, creation of a copy of the single Asp32 residue extracted from the protein and of a copy of the whole protein with all the charges of the Asp32 residue set to zero;

8. the number of grid nodes was determined in such a way that in the *fine grain* step of the focusing procedure the separation between nodes would be between 0.25 Å and 0.50 Å, while the dielectric constant ϵ inside the protein, suggested as 20 in the APBS tutorials, but changed to other values during the calibration phase (see below);

9. APBS run on every input file, each loading a different pqr file and referencing it to its "Asp32" version (or to self);

10. the extraction is best done through scripts finding the relevant information in the standard APBS output files and copying them onto a single new file;

11. and 12. the shift, average and standard deviation can be easily calculated via scripts parsing the file written during the last step.

A calibration phase is needed in order to verify that the chosen dielectric constant ($\epsilon = 20$) is able to provide for NP4 results comparable with those obtained in [67]. The methodology is thus applied to snapshots from an NP4 MD in the closed configuration on Asp30, and the resulting pK_a average shift are $+4.9 \pm 0.5$, which, added to the model 4.0 value for Asp, should lead to a pK_a of 8.9 for Asp30 in the closed form of NP4, thus providing a close match with the reference results obtained from CpHMD calculations.

4.3 Results

4.3.1 Experimental results

By combining ps pump and probe with ns flash-photolysis measaurements performed on the carboxy-E27Q mutant (see [55] for more information about the mutant) it is possible to notice some peculiar trends differentiating *wt*NP7 from the E27Q mutant.

In Figure 4.1 the results are shown, where the empty gap in the 1 ns-10 ns region is due to the fact that the pump and probe measurements can be extended to 1.9 ns for the longest, while the ns technique resolution is no lower than 20 ns.

These peculiar trends that can be extracted from Figure 4.1 are:

1. sub-ns kinetics

   ***wt*NP7** the sub-ns kinetics comprises two kinetic phases, whose amplitudes and apparent rates increase at pH 5.5;

   E27Q also the sub-ns kinetics becomes faster and larger when the pH is 5.5, but the difference is smaller than for the wt protein; this phase can be described with two exponential decay functions.

2. bimolecular rebinding

   ***wt*NP7** the bimolecular rebinding has a remarkable heterogeneity at pH = 7.5, with a fast (100 µs) rebinding followed by slow, ms lived decay independent of CO concentration[1]. At 5.5, the bimolecular phase is much faster but not less heterogeneous (with lifetimes 30 µs and 300 µs);

   E27Q here too there is a slightly more heterogeneous bimolecular rebinding kinetics (three exp decay steps are needed, with lifetimes about 10-50 µs, 200 µs and 2 ms), but the effect of pH is negligible.

[1] This slow CO concentration independent decay might be explained in terms of a different species that is reacting more slowly because of some structural rearrangement, possibly for the entrance of water molecules in the heme pocket and the consequent competitive effects in binding.

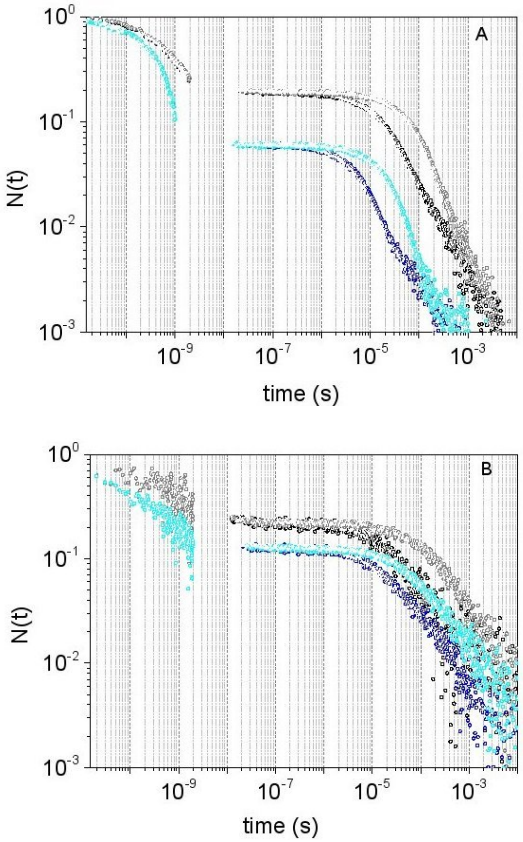


Figure 4.1: Comparison of CO rebinding kinetics of *wt*NP7 (A) and NP7(E27Q) (B). T=20 °C. Blue: pH 5.5, p_{CO}=1 atm. Cyan: pH 5.5, p_{CO}=0.1 atm. Black: pH 7.5, p_{CO}=1 atm. Grey: pH 7.5, p_{CO}=0.1 atm. Unpublished data.

The different behaviour between the *wt* and the E27Q mutant in terms of pH-induced shift shows a definite decrease in the efficiency of the pH switch and a slower rebinding on the whole.

As far as the geminate rebinding phase is concerned, when the pH is lowered ligands find escaping more difficult from the interior of the protein. The increase in the apparent rate suggests that there may be an increase in the rate of the rebinding step (CO binding to Fe). This may be mixed to the expected change in the exit/entry rates due to the conformational consequences of the pH change. Heterogeneity in the fast rebinding phase (ca. 0.5×10^{-10} s and ca. 0.5×10^{-9} s) may mean that there is modulation by temporary docking sites located near the distal pocket.

About the bimolecular phase, the almost complete loss of pH sensitivity can be explained in the following terms:

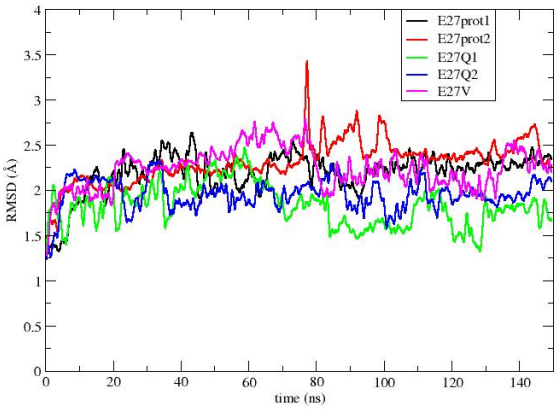


Figure 4.2: RMSD (protein backbone) against final *wt*c (at 300 ns) of the variant structures. (Grace)

1. Glu27's protonation is crucial to trigger the conformational change (i.e. it takes part directly as a primary actor in said conformational transition), so that when Glu is mutated to Gln and the deprotonation cannot happen the conformational change is made impossible, or

2. Asp32, so far understood as the crucial residue for the conformational change, undergoes a pK_a shift due to the different electrostatics generated by the presence within the proximal cavity of Glu27 or Gln27.

As Glu27 is reported to be always charged (see [55]), and as some pH sensitiveness is still visible, hypothesis 2) seems worth investigating.

4.3.2 Simulating the mutants

RMSD and distances

After 150 ns of MD, the backbone RMSD profile along the trajectory is shown in Figure 4.2. During the last 70 ns all forms seem to reach a similar level of stability, for which the particular rigidity of the β-barrel might be credited.

A measure of the resemblances and differences between the structures is given by the cross-RMSD values in Table 4.2. Here the energy-minimized average structures (see below) are overlaid and their RMSD is calculated. Although it is easy to identify which structure resembles most another, a clear division in groups is not available and all five structures seem to be part of a gradually differing single class.

RMSD (Å)	prot2	Q1	Q2	V
prot1	2.203	2.106	1.785	2.438
prot2	-	2.119	2.240	2.207
Q1		-	1.635	1.871
Q2			-	1.821

Table 4.2: Cross-RMSD values for backbone atoms between the energy-minimized average structures, including N- and C- termini. (VMD-RMSDTT, see Section 2.4.1)

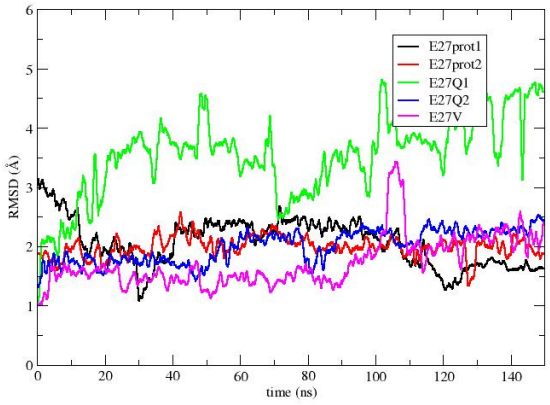


Figure 4.3: RMSD (heme+CO, heavy atoms) against final *wtc* (300 ns) of the variant structures after 150 ns. (Grace)

Figure 4.3 shows the heavy atom RMSD for all the mutants' heme cofactor and CO ligand, upon structure alignment of the protein backbone with the reference's (*wtc*). The average value indicates a common trend for all the simulated systems but E27Q1, which stands out as peculiar in terms of heme positioning. This feature shall be dealt with in what follows.

Average structures

In order to collect average structures, the following trajectory intervals were used, on the basis of the information provided by Figure 4.2:

E27prot1: 140-150 ns

E27prot2: 120-130 ns

E27Q1: 130-140 ns

E27Q2: 130-140 ns

E27V: 120-130 ns

Submitting the average structures to an energy-minimization process is a common practice aimed at obtaining average structures that are physically likely, at least from a steric point of view. This process is described in Appendix A.1.2. In what follows, then, the reference structures will be the energy-minimized average ones taken from the time windows as in the scheme above from each trajectory.

RMSD plots of different backbone selections vs. the energy-minimized average structures are shown in Figure 4.4. As the black line represents the whole backbone, the red excludes the first four residues (N-terminus) and the green line excludes the N-terminus and the AB and GH loops, it is apparent that the largest part of the RMSD variability is due to the motions of the N-term and said loops. Apart from the very last part of the E27prot2 and E27Q2 trajectories, the high degree of stability shown by the green plots (the protein core, where the β-barrel's low contribution is predominant) is comparable between all five trajectories.

The new position of Glu27 in the average structures (Figure 4.5) shows some differences:

- E27 is bent downwards (i.e. occupies the stable *wt* closed position) in E27prot1 and has a transient interaction with Ser137 (see Figure 4.7), while in the Q-mutants, the Ser137/Q27 interaction is more consistent and stable.

- The heme cofactor displays very similar positions between the prot1 and Q2 structures in terms of rotation and plane slanting, while prot2 is similar in terms of rotation but has a slanting that is similar to Q1's which, in turn, displays a peculiar rotation that is responsible for the difference in RMSD in Figure 4.3. This effect on the heme's position, which is probably due to a different tension on the His60-Fe-CO axis at least judging from His60 sidechain's orientation (Figure 4.8) might be a sign of a conformational variability partially explaining the heterogeneous character of the experimental data.

- In spite of such displacement, the ligand cage provided by hydrophobic 123-125-132-135 residues is preserved (Figure 4.6-b).

As stated above, it is especially the heme RMSD plot (Figure 4.3) that marks a difference in the structures between the E27Q1 mutant on one side and the rest on the other.

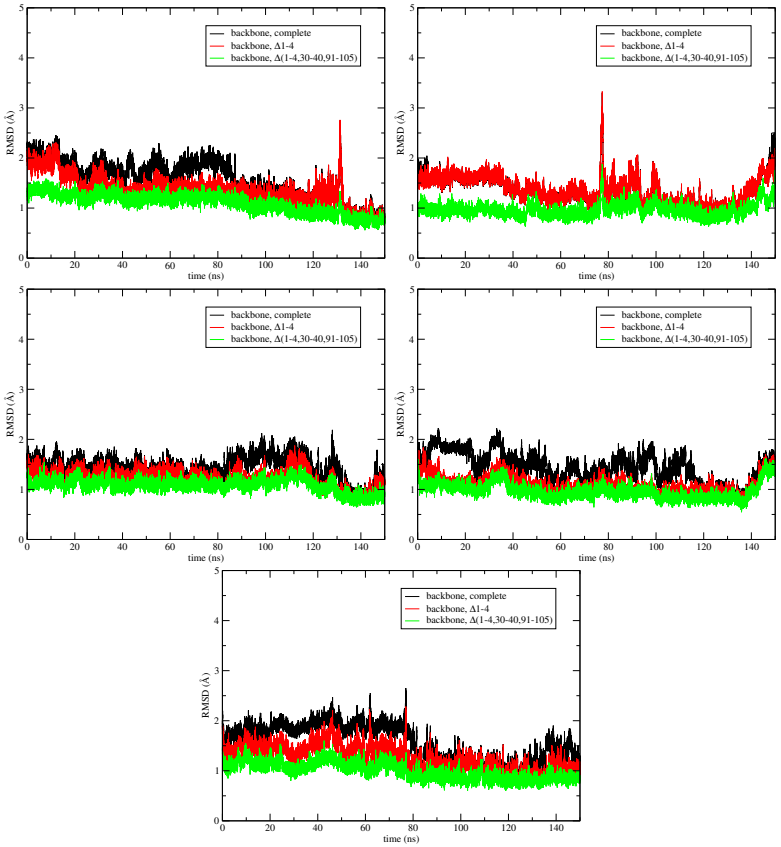
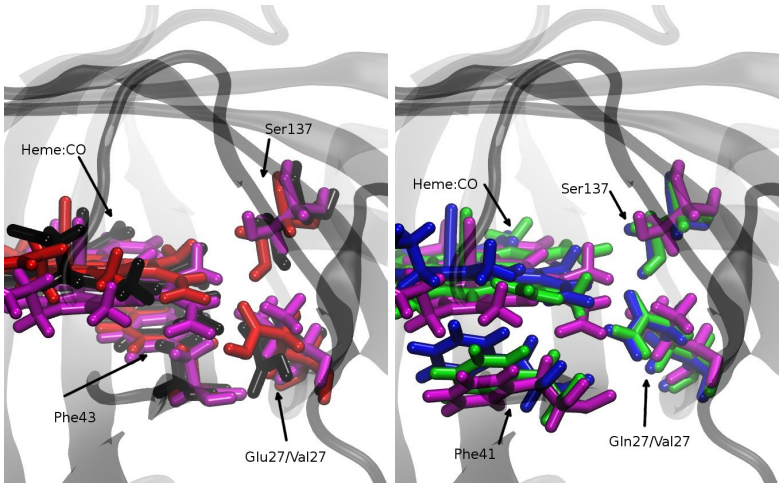


Figure 4.4: RMSD (black: protein complete backbone, red: backbone without the N-terminus LPGE-, green: backbone without the N-terminus LPGE- and the AB and GH loops) against their respective energy-minimized average structures after 150 ns. Above: E27prot1/2 (left/right). Centre: E27Q1/2 (left/right). Below: E27V. (Grace)

This division, though, is not absolute, also as observed above: e.g. the cross-RMSD values (Table 4.2) for the average structures indicate that Q1 and Q2 are anyway structurally more similar between themselves than the two protonated forms between themselves.

The new position of Glx27 is compatible with a H-bond between one of the sidechain oxygen atoms (even the one hosting the proton, or the only surviving the amidation to Gln) and the hydroxyl of Ser137. The other oxygen atom or the sidechain aminic nitrogen in Gln, can form bridges with the backbone around residues 41-43. On the contrary, Val27's sidechain cannot obviously form H-bonds or salt bridges with neighbouring residues and apparently chooses a slanted, least-clash orientation between Ser137 and the heme. The distances are shown in Figure 4.7.



(a) (b)

Figure 4.5: Glx27's position relative to the heme in the energy-minimized average structures in E27prot1 (black), E27prot2 (red), E27Q1 (green), E27Q2 (blue) and E27V (magenta) upon backbone alignment. Ser137 and Phe43 (a) / Phe41 (b) are shown. (VMD)

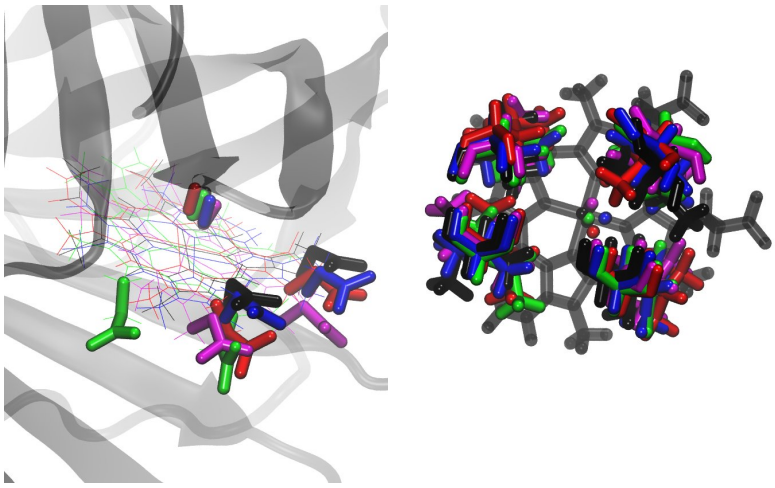
As can be seen, all energy-minimized average structures (Figure 4.5) have Glx27 stretched inside the proximal cavity, except E27prot1. All the same, inspecting the final structures and checking the distance plot in Figure 4.7, the bond with Ser137 seems to be much less permanent than the average structures lead to imagine, and eventually it is only the E27Q1 structure, with its otherwise peculiar structure, the one where no breaking of such bond is reported, while all others (E27V excluded for obvious reasons) experience, for shorter or longer times, some return to the old original form with interactions with Thr168 and Tyr175.

His60's position

Another interesting element is the position of His60. Besides its stretching in the mutants, the imidazole ring exhibits different positions along the trajectories, especially in Q1, as visible in the energy-minimized average structures in Figure 4.8.

Plotting the distributions of the dihedrals between planes CA-CB-CG and CB-CG-CD2 of His60 (Figure 4.9), some interesting features include:

- E27prot1/2 and E27Q2 have very similar profiles with a primary (around 190°) peak and a secondary one (around 120°)
- E27Q1 has a peculiar profile, with the primary peak shifted to around



(a) (b)

Figure 4.6: (a) Heme's position relative to the heme in the energy-minimized average structures in E27prot1 (black), E27prot2 (red), E27Q1 (green), E27Q2 (blue) and E27V (magenta) upon backbone alignment. (b) Hydrophobic cage and CO ligand in the energy-minimized average structures (same colour coding). (VMD)

240°and a much lower secondary peak similar to the others'

- E27V has a profile similar to the E27Q1 one, with a much more prominent secondary peak shifted at about 150°.

The data collected so far can be then summarized as follows:

- the general three-dimensional fold of the protein is not affected by the mutations/variations

- there are local effects, concerning the longest loops but also and more importantly the heme cavity, where the heme itself and His60, besides all the residues directly interacting with residue 27, seem to feel the influence of the mutation/variation

- the full extent of the influence over the heme, in particular, and on the Fe-ligand bond is not attainable through classical methods; these methods, though, are capable of highlighting a definitely different stress that these mutations/variations cause to other elements of the protein fold, e.g. to the β-barrel core, and to the histidine-cofactor-ligand subsystem.

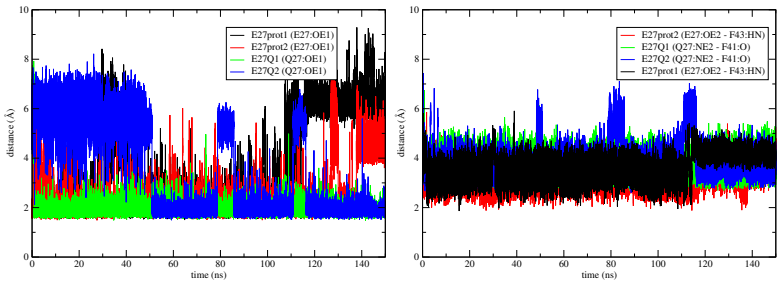


Figure 4.7: Distance between one of the sidechain oxygen atoms of Glx27 and Ser137:HG1 (left), and between the other sidechain oxy/nitrogen atom and the backbone in residue Phe43/41 (right). (Grace)

	back tunnel	cage[a]-to-back	cage[a]-to-front	cage[a]-to-top
*wt*c s1	80%	<10%	28%	50%
*wt*c s2	34%	<10%	26%	55%
E27prot1	46%	<10%	23%	31%
E27prot2	47%	10%	27%	24%
E27Q1	25%	<10%	13%	<10%
E27Q2	41%	13%	39%	53%
E27V	29%	<10%	41%	59%
*wt*o	27%	<10%	100%	0%

Table 4.3: MDpocket frequencies for tunnels. [a] For "cage" the hydrophobic cage is meant, which is completely disrupted by Asp32 deprotonation's effects in *wt*o; in this case the proximity of the CO ligand is considered. See Figure 4.10.

Fpocket and ILS

In Table 4.3 the frequencies of tunnels are specified for all the MDpocket runs. For the mutants, the sampled trajectory windows are those used to calculate the average structures.

The other structures follow considerations exposed in previous reports (see 4.3.2). For completeness, they are listed below.

***wt*c s1:** 180-190 ns

***wt*c s2:** 220-230 ns

***wt*o:** 280-290 ns

The results allow us to disclose a few trends:

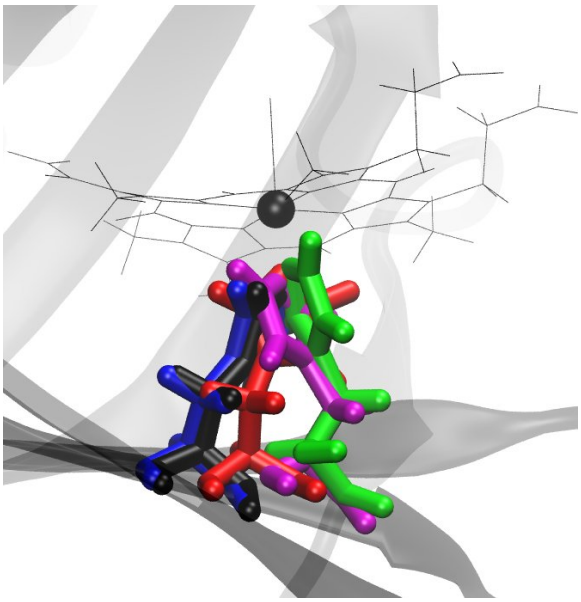


Figure 4.8: Position of His60 (upon heme alignment) in E27prot1 (black), E27prot2 (red), E27Q1 (green), E27Q2 (blue) and E27V (magenta) in the energy-minimized average structures. (VMD)

a) access to the back tunnel seems easier in the closed form of *wt*NP7 than in the E27Q mutant. Thus, in the former, depending on the specific window, the access can be as large as 80%, whereas in the E27Q mutant is found to range from 25% to 41%.

b) the exit to the bulk solvent though the cage-to-top pathway is similar in the *wt*NP7 and in the E27Q mutant.

These trends suggest that the ligand, upon flash photolysis, could likely rebind from regions proximal to the the distal cavity (i.e. site 1) and from more distant regions (i.e., site 3), thus explaining the two kinetic phases in the geminate rebinding. However, since the access to site 3 seems to be more feasible for wt Np7, the ligand should be retained more effectively in the *wt* protein, thus leading to a larger fraction of geminate rebinding, as found in the experimental assays.

ILS runs on the considered trajectory segments offer the values in Tables 4.4 and 4.5, in terms of appearance of docking sites 1a, 1b, 2, 3 (Figure 4.11) and others. An additional 1c (Figure 4.11) and a 4th docking sites (Figure 4.12) are defined in the distal pocket (columns “1c” and “4”). “BT” (Figure 4.12) stands for “back tunnel” as defined in Table 4.3; “SED” (Figure 4.12) and “SEA” (Figure 4.13) identify the “side exposure”, meaning the ligand is almost touched by the ILS isocontour of an external intrusion from side/top

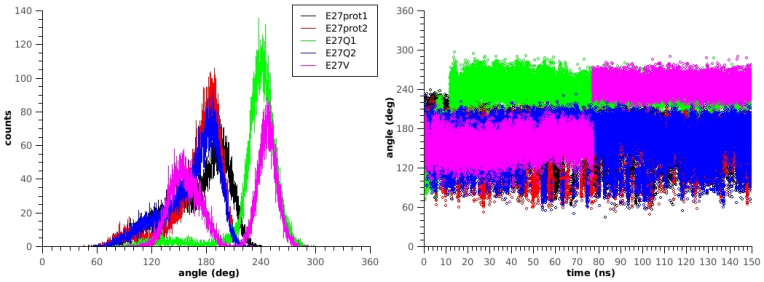


Figure 4.9: His60:CA-CB-CG-CD2 dihedral angle plot. Left: angle distribution. Right: time-vs-angle. (Qtiplot)

	1a	1b	1c	2	3	4
wtc s1	-3.9	+2.4	+14	-5.8	-4.7	+8.0
wtc s2	-1.9	+3.0	+14	-4.2	-5.9	+7.1
E27prot1	-4.0	-2.2	+3.1	+1.9	-5.3	+6.7
E27prot2	-4.5	-1.3	-0.3	-2.6	-6.5	+1.7
E27Q1	-2.8	+0.9	12	-3.5	-5.4	+3.1
E27Q2	-5.2	-1.9	-	-0.1	-4.9	+0.9
E27V	-0.3	-1.9	+1.6	-0.1	-5.9	+0.7
wto	-6.4	-1.8	-3.2	-1.0	-5.6	+0.3

Table 4.4: ILS isovalues for the docking sites of different structures.

behind the D-propionate's side and behind the A-propionate's side respectively; "CTB" (Figure 4.13) stands for "cage to back" as already seen in Table 4.3; "BTS" (Figure 4.14) means a direct connection between the back tunnel and the side (behind the D-propionate's side) and finally "BTF" (Figure 4.14) is a direct connection between the back tunnel and the barrel's mouth through the proximal pocket or on the heme's rim at the A- propionate's side.

One point is important enough to be underlined here. In the "open" structure (*wto*), once the back tunnel is connected (the bottleneck is between any of the sites labelled as 1 and site 2), it is automatically connected to the cage and to the side, while at lower energy isovalues, sites 2-3 form one single cavity, and the distal cage, is connected to the sites 1 and to the side (next to the D propionate).

Taken together, the ILS and Fpocket data about cavities and tunnels do not allow to point out significant qualitative differences in the cavity/tunnel system of the mutants, once taken into account the different position of the heme that characterizes the E27Q1 structure, which influences some of the energetic barriers, but which also seems only one of the possible conformational variants

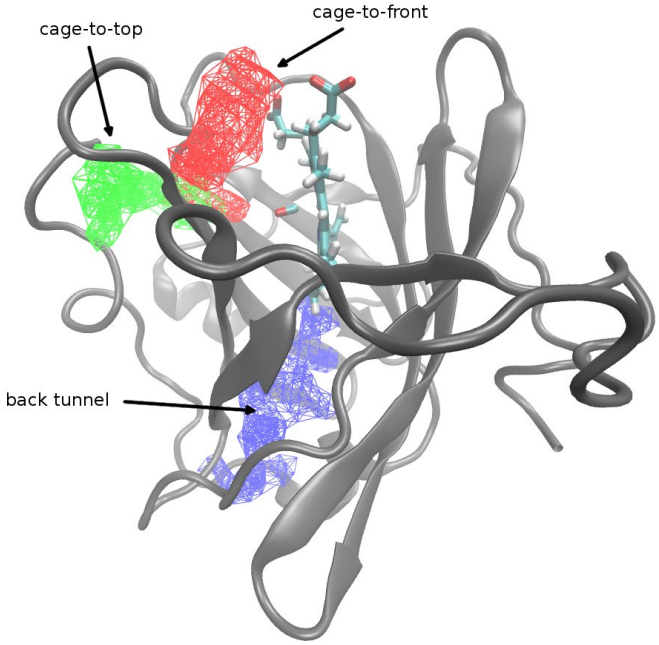


Figure 4.10: Tunnels in NP7 mutants. Supporting image: *wt*c s1; back tunnel and cage-to-top images caught at iso 50% frequency, cage-to-front at iso 28% frequency (see Table 4.3). (VMD)

of a configuration-shifting structure.

One important remark, though, is possible by noticing in Table 4.4 that the energetic minima 1a and 2 for E27V are among the highest, while minimum 4 it is one of the lowest. This agrees with the observations in 3.3.5 about the cavity system of NP4 (which, like all other NPs, has a valine instead of NP7's Glu27), and shows the importance of Glu27's steric influence on the cavity system.

The results in Table 4.5 show that the most favourable pocket is site 3, which appears at an energy value almost constant for all proteins. Thus, if a photolyzed ligand gets site 3, it will show a slower geminate rebinding due to the energetic stability of the pocket in the rear of the tunnel. It is also worth noting that transition from site 1 to site 2 is favourable in the *wt* protein, but that the opposite trend is found for the E27Q mutant (note also the larger barrier for motion through the back tunnel in Table 4.4 for the E27Q mutant). This suggests that a photolyzed ligand will have larger tendency to occupy site 3 in the *wt* protein than in the E27Q mutant, thus explaining the larger fraction of geminate rebinding in the former.

	BT	SED	SEA	CTB	BTS	BTF
wtc s1	+2.6	+10[a]	+19	+60	+60	+29
wtc s2	+5.3	+5.4[a]	+18	+50	+53	+30
E27prot1	+5.2	+12	+35	+59	+31	+60
E27prot2	-0.1	+3.6	+10	+27	+18	+65
E27Q1	+3.2	+11	+30	+34	+45	+24
E27Q2	+9.1	+0.4	+20	+33	+24	+55
E27V	+8.2	+4.0[a]	+19	+32	+36	+53
wto	+10	-	+11	+10	+10	+50

Table 4.5: ILS isovalues for the tunnels of different structures. [a] "Top" connected at lower isovalues. For the tunnels definitions, see 4.3.2.

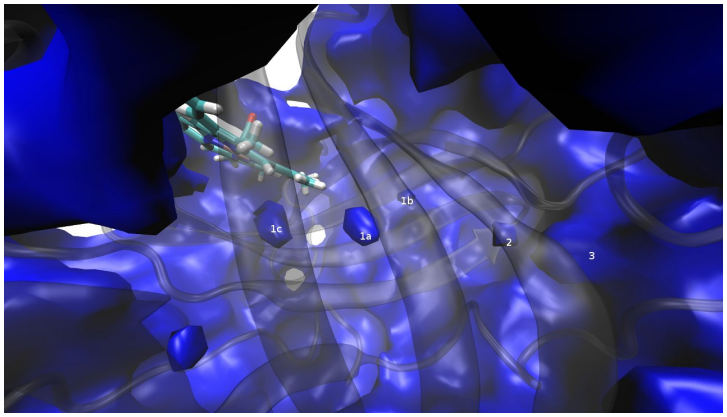


Figure 4.11: Cavities in the interior of the protein. (see Table 4.4). (VMD)

4.3.3 APBS results and how they can be interpreted

Due to the presence of many positively charged residues in the far end (with respect to the heme pocket) of the barrel of NP7, its isoelectric point is much higher than that of all other NPs (see Table 1.2). Globally then, for NP7 the creation of a negative charge by means of the acidic dissociation of a ionizable residue is favoured, and the pK_a for Asp32 in NP7 should be expected to be lower than the pK_a value of Asp30 in NP4. The E27Q mutation amplifies this reduction, as it increases the charge imbalance and makes the dissociation of Asp32 even more favourable: therefore the E27Q pK_a for Asp32 is expected to be even lower than in *wt*NP7.

The results of the APBS run on *wt* NP7c and NP7o, NP7(E27p), NP7(E27Q) and NP7(E27V) are summarized in Table 4.6.

From these figures, a prediction for the experimental pH of conformational

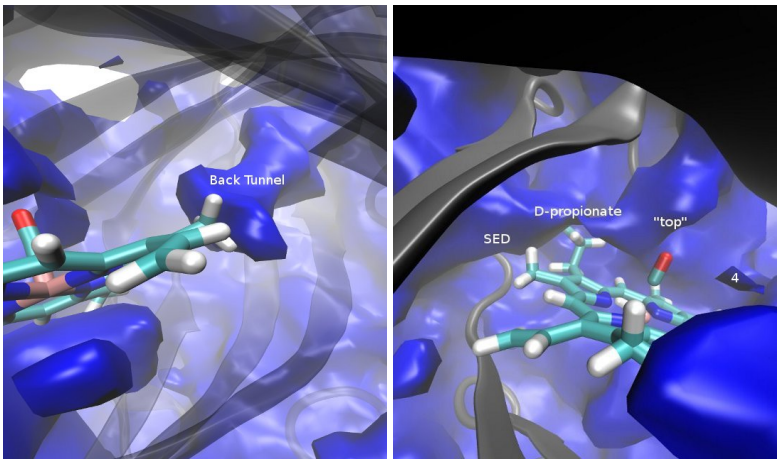


Figure 4.12: Back Tunnel (BT) and Side Exposure by D-propionate (SED), with cavity 4. (see Table 4.4). (VMD)

change for the E27Q mutant can be done in a rough way as in Table 4.7. The result is of course at least partially arbitrary, lacking the methodologically validating procedure (CpHMD) followed for NP4 in [67], and a difference in a decimal can lead to important differences in the closed:open ratio for any structure. All the same, the trend is rather reliable and worth noticing.

First of all, the data about NP4 can fit the experimental observation in [66] of a heterogeneous behaviour of NP4 at pH 7.5. The 1:10 and 10:1 ratios are mirrorlike because the apparent pK_a is 6.3, right in the middle between the experimental pH values (the experimental apparent pK_a is 6.4, which also agrees with the estimated pK_a for closed and open forms from CpHMD calculations: 8.5 and 4.3, respectively). Assuming that the open form, being more flexible, can assume a variety of actual configurations, with water molecules able to invade the heme cavity and therefore with different competition and evacuation patterns on ligand rebinding, the above mentioned heterogeneous behaviour can be justified.

The estimates in Table 4.7 indicate that in the *wt* case at pH = 5.5 there is an almost even open-closed mixture. Then, if the ligand rebinds to the closed form of the protein, where the cage must be preformed and the water molecules that fill the cavity in the open species should have been already released, at pH 5.5 CO should exhibit a faster bimolecular rebinding than at pH 7.5, where the open form will be the predominant species. In contrast, since the apparent pK_a of the E27Q mutant is 4.9, the fraction of open species must be predominat at the two pHs, thus explaining the similar profiles obtained for the mutant.

In any case, the finding that the pK_a appears to be much lower for E27Q explains in a convincing way the mutant's lack of sensitivity to pH and the fact that its bimolecular rebinding signals are much more similar to *wt*'s at neutral

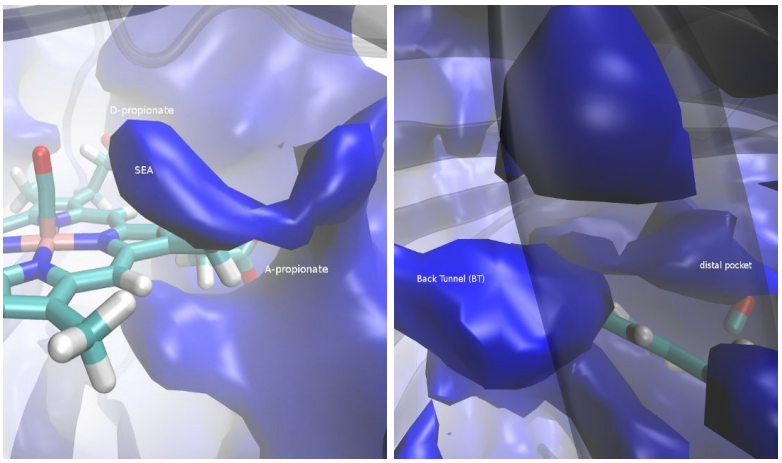


Figure 4.13: Side Exposure by A-propionate (SEA) and Cage-To-Back (CTB). (see Table 4.4). (VMD)

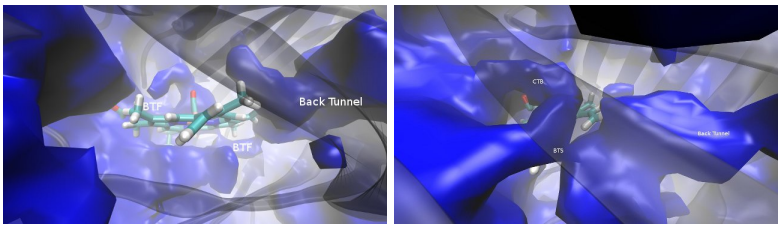


Figure 4.14: Back-To-Front (BTF) and Back-To-Side connections (BTS). (see Table 4.4). (VMD)

pH.

4.4 Conclusions

The key role of Glu27 in NP7, and therefore of the unique NP family-wise mutation Val→Glu seems to get a clearer contour.

As the steric presence of Glu27 is to be held responsible for the particular heme orientation selectivity that NP7 displays, now to its charge the functioning of Asp32's pH switch between the two physiologically relevant pH values can be ascribed, at least partially.

In fact, in comparison with other members of the NP family, the presence of a higher number of positively charged Lysine residues in the back of the protein would considerably lower the pK_as of all negatively charged residues, including Asp32's, and this would lead in turn to a shift of the pH window where the protein undergoes a conformational change between the closed and open

structure	traj. avg.	struct. avg.
wtc1	+3.0±0.3	+2.9±0.4
wtc2	+2.7±0.4	
E27p1	+2.9±0.3	+2.7±0.4
E27p2	+2.4±0.3	
E27V		+2.3±0.3
E27Q1	+2.0±0.3	+1.9±0.5
E27Q2	+1.7±0.6	
wto1	+0.0±0.4	+0.0±0.3
wto2	+0.1±0.3	

Table 4.6: Asp32 pK_a shifts. (APBS)

structure	c/o pK_a	apparent pK_a	c:o ratio pH 7.5	c:o ratio pH 5.5
NP4	8.9/4.0	6.3	1:10	10:1
*wt*NP7	6.9/4.0	5.4	1:125	1:1.25
NP7(E27p)	6.7/4.0	5.3	1:167	1:1.67
NP7(E27V)	6.3/4.0	5.1	1:250	1:2.5
NP7(E27Q)	5.9/4.0	4.9	1:400	1:4

Table 4.7: Projection of apparent pK_as for NP7.

forms, or, equivalently, to a less efficient pH switch mechanism between insect's saliva and victim's blood due to an incomplete conformational transformation.

The Val→Glu mutation, introducing an extra negative charge in NP7, only partially counterbalances the charge difference, but apparently through its interaction network is able to modify the electrostatics in the heme's environment in such a way as to keep Asp32's acidity constant high enough as to allow the pH switch to work between the two physiological pH values.

This mutation also seems to cause a modification in the structure of tunnels. The presence of a better defined tunnel in *wt* may explain the larger amplitude of the second geminate phase (ca. 0.5×10^{-9} s), whereas in E27Q this phase has approximately the same amplitude as the faster one (ca. 0.5×10^{-10} s). In *wt*NP7, the more accessible secondary docking site(s) may then be held responsible for the larger amplitude of the ca. $0.5\,\mathrm{s} \times 10^{-9}$ s decay. The computational results then offer a good agreement with the experimentally observed results for the geminate rebinding phase. The difference in the extent and the rates between pH 7.5 and 5.5 is much larger in *wt*NP7 than in NP7(E27Q): the data for the mutant are both more similar to those at neutral pH for the *wt*, suggesting that the effect of the imposed pH change is weaker for the former. For the *wt* protein the rates and the amplitudes of the exponential decays become larger

when pH is lowered, and the corresponding increase (in rates and amplitudes) observed in E27Q is much less pronounced.

Besides adding experimental data about the NP7(E27Q) and NP7(E27V) mutants in comparison with *wt*NP7, further computational studies must be developed in order to better understand whether Glu27's role is accompanied by other unique residues' concurring action, and in order to refine the model, possibly using the same CpHMD methods already cited above, thus seeking a better agreement between the model itself and the experimental data.

Chapter 5

NP7 and membranes

5.1 The questions addressed

NP7 is known since 2003 [30, 32] and almost immediately its feature of binding to anionic phospholipidic membranes was recognised [33].

The connection between the family-unique polybasic sequence found in NP7 and the possible adsorption to negatively charged lipidic layers was advanced in 2004 [31].

Deeper insight about the membrane-binding properties was offered by [57, 58], and, more recently, by [70].

Suggestions contained in [70] and personal communications with the authors led to focus the attention on residues Arg141, Ala145, Pro147 and Ala154 as those among the NMR-labelled ones giving different signals between the in-solution and membrane-bound situations.

In the light of the results shown in the previous two chapters of MD studies of NP7 in solution, and especially in the light of the existence of an axial tunnel connecting the distal cavity to the "back" of the protein, affected by the pH conformational switch and with possible implications on the NO load-and-release mechanism, a computational study was performed with the aim to find detailed clues as to how the membrane binding happens, what its structural impact is, and what signs of its features can be detected experimentally.

5.2 Methods

5.2.1 The force field

The choice of a force field is obviously crucial for the quality of the results. The recent publication of the new, experimentally validated CHARMM36 parameters for lipids [112], oriented the choice of the force field for representing NP7 interacting with a model membrane.

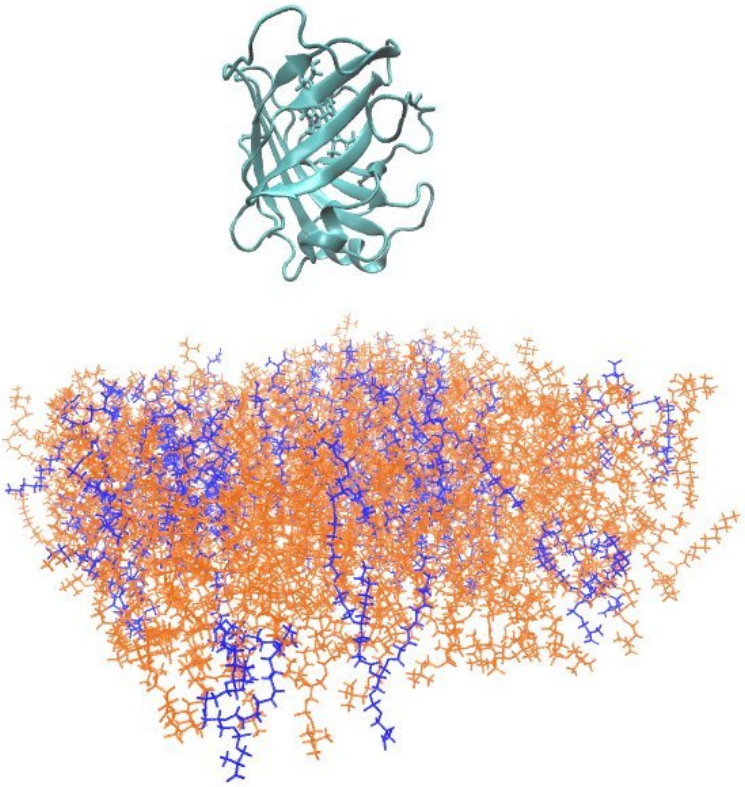


Figure 5.1: Carmen, the system as it looked before MD runs (water molecules are not shown). (VMD)

5.2.2 Building the systems

MD simulations were run on two different systems, composed each of a phospholipidic bilayer and a NP7 molecule, in carboxy closed and open forms. The two systems, (approximately 110k atoms each) were named “Carmen” (“Chiusa” - closed NP7 + membrane), shown in Figure 5.1, and “Azucena” (“Aperta” - open NP7 + membrane).

The two systems were built using the same model membrane and adding to it the two protein variants.

The protein

For the proteic part, two snapshots were chosen from the stability regions of the trajectories described in Chapter 3. The snapshots were stripped of all water molecules and counterions and then merged (VMD script) with the membrane part of the system.

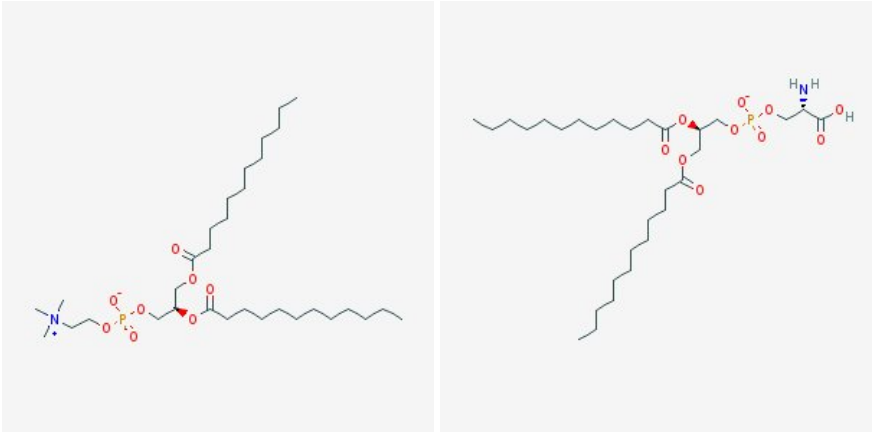


Figure 5.2: Dilauroyl-L-α-phosphatidylcholine (DLPC - left) and dilauroyl-L-α-phosphatidyl-L-serine (DLPS - right). (Source: NCBI Pubchem)

The membrane

The membrane was built using the Membrane Builder [113, 114] input generator on CHARMM-GUI [115].

For the membrane composition, according to [57], was chosen as a 3:1 mix of L-α-phosphatidylcholine (PC) and L-α-phosphatidyl-L-serine (PS) for the polar heads, while for the lipidic tails two lauroyl chains (DL) were chosen[1]. The two phospholipidic molecules (DLPC, DLPS), are shown in Figure 5.2.

A square 85 Å×85 Å bilayer was built using 96:32 molecules of DLPC:DLPS per side. This was enclosed in a 85 Å×85 Å×140 Å tetragonal water box, with the bilayer stretching in the $z = 0$ plane and dividing the water box in two non-communicating volumes, one "above" and one "below" the bilayer (positive and negative z coordinates, respectively). In terms of the z dimensions, the bilayer's thickness, head to head, measured about 20 Å, leaving two volumes about 60 Å deep on both sides for water and ions (and for the insertion of the protein)[2].

[1] Lauroyl tails were chosen because their chains, shorter than those from other more physiologically relevant fatty acids, allow a smaller system in terms of number of atoms, and therefore represent, with suitable adjustments, a fair compromise between computational complexity and model realism. Just as an example, the use of palmitoyls instead of lauroyls on the same membrane system would have added over 6k fully interacting atoms to the atom count. The drawback of a less massive membrane is anyway overcome by positional constraints on the lipidic tails imposed to prevent the membrane from being displaced by the interaction with the protein. An important plus in the choice of DLPC was the fact that this particular lipid is one of the six used in [112] to validate the CHARMM36 force field, and the computational performance of a model of DLPC based on such force field parameters is therefore to be considered of high quality.

[2] This thickness was chosen so that the protein could fit conveniently in the top water volume without being too close to the nearest surface and allowing for several layers of water molecules in between, and at the same time being far enough from the negatively charged bottom layer of

value	all	s1	s2	s3	s4	s5	s6
cgem[a] (×1000)		10					
timesteps (×1000)		25	25	25	100	100	100
timestep value (fs)		1	1	1	2	2	2
Langevin[b] temp	on						
Langevin[b] damping	10						
Langevin[b] H	no						
Langevin[b] piston (LP)		off	off	on	on	on	on
LP[b] period				50	50	50	50
LP[b] decay				25	25	25	25

Table 5.1: NAMD settings for membrane equilibration. [a] Conjugate gradient energy minimization steps. [b] For Langevin dynamics, see reference settings in Table A.2.

At this point, the atom count read 26688 atoms for the bilayer, about 87300 atoms belonging to TIP3 water molecules, 107 K^+ and 43 Cl^- counterions to balance the overall negative charge of the DLPS molecules (each carrying a -1 charge), compatible with a 0.15 M KCl concentration.

The CHARMM-GUI online server was used to assemble the bilayer. Its output includes the NAMD input files for a 6-step equilibration phase and the prototype input file for production dynamics. Therefore, unlike all other protein simulations mentioned so far, for the system comprising the membrane alone the settings were not based on those shown in Appendix A, but as in Table 5.1.

The membrane-only system therefore underwent an initial 10000 steps energy minimization procedure and then an equilibration phase simulating an overall period of 675 ps. During this equilibration steps collective variables (colvars) were set and used in order to constrain and monitor the membrane thickness and area, while other extra constraints were applied on improper dihedrals in order to maintain the planarity of the membrane.

This preparatory phase was followed by a full production MD run of the membrane alone spanning 50 ns and set as the last equilibration step.

During these steps the area-per-lipid and the layer thickness were constantly monitored and this showed that the system was able to retain the stability gained during the equilibration process, as can be seen in Figures 5.3 and 5.4.

Only at this point were the membrane and the two protein structures assembled in two separate systems.

the nearest periodic image of the membrane so that it would not “feel” its influence.

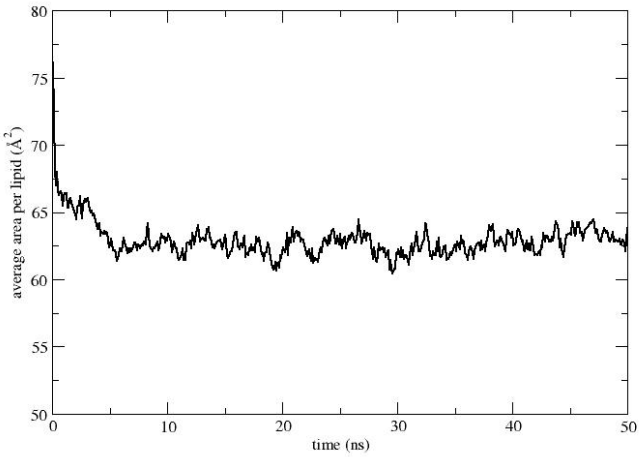


Figure 5.3: Area-per-lipid (top and bottom layers share the same values) for membrane in free MD. (Grace)

5.2.3 Assembling the systems

The Carmen and Azucena systems were assembled by placing the protein structure (derived, it is worth remembering, from snapshots extracted from stable parts of trajectories of the same proteins in solution, i.e. with the protein already well equilibrated and in reliable simulation conditions) into the membrane's large water box.

In order to do so, a specific VMD script allows to overlap the two pdb files, including the protein in the water box and deleting from it all water molecules and ions that happen to be in the position where the protein is set.

The protein was set on the "upper" side, that is where $z > 0$, at a minimum distance from the polar heads of the membrane of 10 Å, the β-barrel axis perpendicular to the membrane's surface and the "back" of the barrel closest to the surface.

Placing the protein at such distance, i.e. where the electrostatic interactions between the positively charged lysines of the "back" of the protein and the negatively charged PS heads are weak yet present, allows to set a Steered Molecular Dynamics run with a constant velocity approach of the protein to the molecule that permits a gradual evacuation of water molecules from in between, a "soft landing" of the protein onto the membrane with a progressive engagement of all the lysines that can be involved in the process.

The reason why the "back" of the barrel was set facing the membrane, as opposed to its "mouth" or other parts, is that it is exactly in the "back" area that

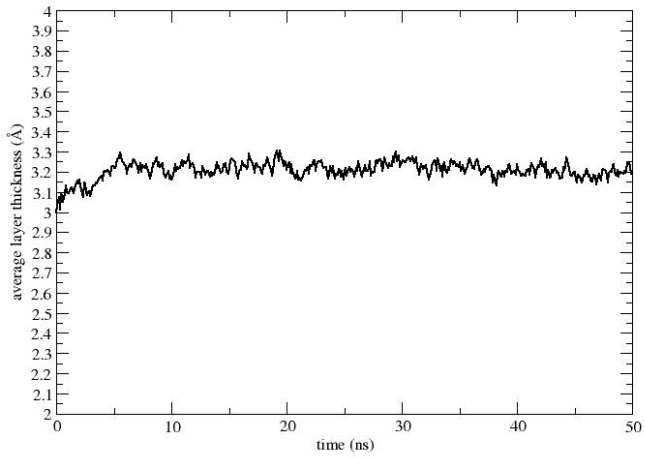


Figure 5.4: Layer thickness (averaged over the two layers and over time) for membrane in free MD. (Grace)

the polybasic sequence that stands out as unique in the NP family is located. The basic assumption is that NP7's peculiar behaviour with anionic membranes is directly related to the peculiar structural feature of an accumulation of positive charge at the rear end of the β-barrel, and that this direct relation is interpretable as an electrostatic interaction with the negative charges in the membrane.

The two systems so obtained underwent the usual energy minimization procedure described in A.1.2. The equilibration phase for the two systems was longer and more articulated than in Table A.1. Its outlines are contained in Table 5.2. After these four equilibration steps, amounting to a 500 ps the systems were deemed ready for the next passage, i.e. SMD, in order to slowly bring the protein closer to the membrane and cause the interaction to gradually activate.

5.2.4 Steered Molecular Dynamics

Steered Molecular Dynamics (SMD) is a MD variant where external forces are introduced in order to drive molecules or components in particular movements of interest. Whereas this technique is often combined with trajectory sampling (see [116]), in the present case its choice was based on the considerations that a free MD would potentially lead either to a long wait before the protein starts moving towards the membrane, a waiting during which the protein itself might rotate and present itself with orientation that would not favour the interaction,

value	all	s1	s2	s3	s4
temperature (K)		$100 \overset{50}{\rightarrow} 300$	300	300	300
Langevin temperature control		no	no	no	yes
Langevin pressure control		no	no	no	yes
steps ($\times 1000$)		25	12.5	12.5	200
timestep (fs)	2				
restraints[a]	yes				
force const.[a] ($\mathrm{kcal\,mol^{-1}\,\AA^{-2}}$)		25	12.5	5	5

Table 5.2: NAMD equilibration settings for Carmen and Azucena. [a]Elastic restraining force applied to the lipidic tails, with elastic constant given in the "force const." row. See reference settings in Table A.2.

or to an immediate free fall towards the membrane, which might prove disastrous for the membrane itself because of the unmitigated impact. A driving force applied to the protein's centre-of-mass could avoid both problems.

NAMD allows to do this in different ways. In the present case, a constant velocity was needed in order to start the protein approach to the membrane at first and then to slow down its electrostatics-driven descent on it without harsh collisions that might disrupt the membrane (or, something that would be a bit less likely because of the rigidity of the protein's β-barrel, the protein fold). The result was the application of a variable force of suitable direction to the protein's centre-of-mass.

Together with SMD, positional constraints were applied in certain simulation steps to the lipidic tails at the centre of the membrane in order to prevent the membrane itself from moving from its z-central position because of the attractive interaction or the collision with the protein.

A constant velocity along a vector that would bring the protein more or less to the centre of the square bilayer once in proximity of its surface (during equilibration the protein was free to drift in the solvent) and whose z-component would amount to $0.1\,\mathrm{\AA\,ns^{-1}}$ was then set for the SMD run. Such a force was applied as a harmonic force with an elastic constant of $5.0\,\mathrm{kcal\,mol^{-1}\,\AA^{-2}}$ on the COM according to its distance from its kinematically projected position along the direction vector with the given velocity magnitude after the simulation time.

The SMD simulation was constantly monitored through the output values of force components and COM position that allowed to detect when the protein had reached the minimum distance and started bouncing on the membrane, i.e. the moment to switch off the steering force. The switch-off was performed by setting the force constant to zero instead of switching the SMD engine off entirely, in order to keep NAMD reporting about COM's position (see Figure 5.5

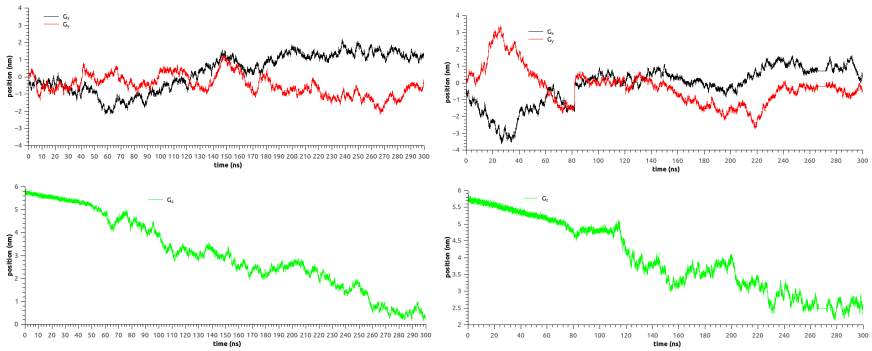


Figure 5.5: SMD monitoring plot for Carmen (left) and Azucena (right) after 300 ns. Above: x and y COM coordinates. Below: z COM coordinate. Applied force mainly along the z axis, off after 50 ns for Carmen, 84 ns for Azucena. (Qtiplot)

as an example).

For Carmen the switch-off could be performed after 50 ns for the steering force and after 60 ns for the lipidic tail constraints. Azucena found more difficulties in settling on the surface of the membrane, and it took 84 ns before the steering force could be switched off, and 110 ns before the lipidic tail constraints could be removed.

5.2.5 Free ligand MD

Free ligand MD simulations were run on Carmen and Azucena, as reported in 3.2.2 and 3.3.4. The differences with the ones in Chapter 3 are the following:

- only initial positions A and B (see Table 3.5 and Figure 3.5), i.e. sites 3 and 1 respectively, were used, as the connectivity of the axial tunnel was the issue to be analysed;
- a larger sampling (10 runs instead of 5) was chosen for the same timespan of 20 ns each;
- the initial structures were the final snapshots (after 400 ns) of the Azucena and Carmen trajectories.

5.3 Results and discussion

5.3.1 After the SMD runs

The production MD phase followed the preparatory SMD runs. The systems obtained at the end of the preliminary phase were considered satisfactory under

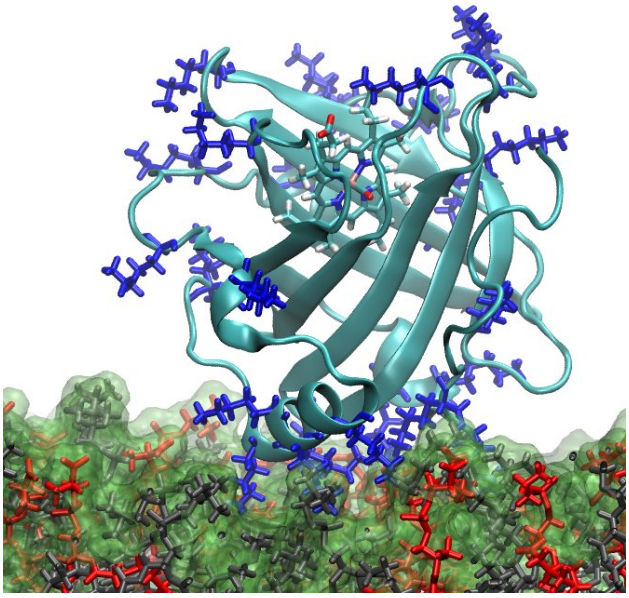


Figure 5.6: Carmen during MD. In blue: Lysines. In grey: PC. In red: PS. (VMD)

several aspects.

First of all, the constant velocity SMD process resulted in the application of a braking force when the proteins accelerated under the coulombic force beyond the target value. This is witnessed by the lower plots in Figure 5.5. Visual inspection shows that this phase had no disruptive or deformation effects on the membranes, thanks to the positional constraints applied to the lipidic tails which kept the membrane in position. Eventually, after the release of such constraints in ordinary MD, the kinetic energy added to the system in the SMD phase led to a definite displacement of the membrane towards $z < 0$, as can be seen in Figure 5.7, while the vicinity of the protein bent the bilayer, in part by colliding and in part for electrostatic attractions between parts of the membrane and of the protein not directly in contact.

Secondly, when the lipidic tail constraint was released, a reorganization of the DLPS molecules in the upper layer of the membrane started, which resulted in a higher concentration of negative charges around the protein contact area, as shown in Figure 5.8. This is worth mentioning: it is an expected effect of the presence of a concentration of positive charge near the surface of an anionic membrane. Of course the light mass of the chosen phospholipids influences this effect, probably mainly in terms of its characteristic times. On the other hand might be a realistic effect with a possible physiological relevance, since the negative charges may have an important role in the docking of other proteins

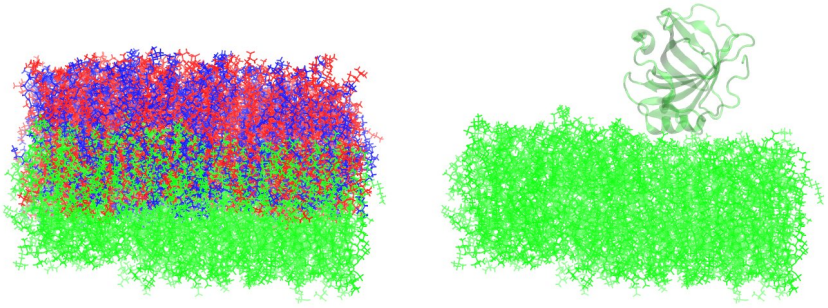


Figure 5.7: Left: side section of the membrane bilayer at the end of the SMD phase for Carmen (red) and Azucena (blue), and Carmen after 192 ns, during the ordinary MD phase. Right: side section of the membrane bilayer of Carmen after 192 ns with NP7; the membrane is slightly bent under the protein. (VMD)

related to the coagulation process, and therefore their concentration around NP7 on the activated platelet's surface might be a disturbing factor in that same coagulation process.

In the third place, the docking approach took place in a significantly longer time for Azucena than for Carmen. The reason for this can be partly pinned in large measure on the conformational variability of the long loops that characterizes the open form of NP7 and that is likely to influence the more or less orderly way the protein approaches the membrane. At the SMD switch-off, the protein in Azucena was rebounding on the membrane with larger amplitudes than in Carmen. This is the main reason why at the same instant Carmen presented a protein maturely docked to the membrane, while Azucena took a longer adjustment time before it could be recognised as being that maturely docked. Despite all this, though, the protein's positioning, the involved residues, the position of the secondary structures, mainly the α-helix, most involved in the interaction, are the same in the two systems, as can be seen in Figure 5.9.

5.3.2 Comparison between Carmen/Azucena and *wtc/o*

After reaching a stable distance between the membrane and the protein back (see Figure 5.6), with the involvement of all available Lysines (12, 13, 19, 116, 149, 152, 153, 157) the production MD run was extended to 400 ns for both Carmen and Azucena.

Comparing the final snapshots (Figure 5.10) and calculating the RMSD for

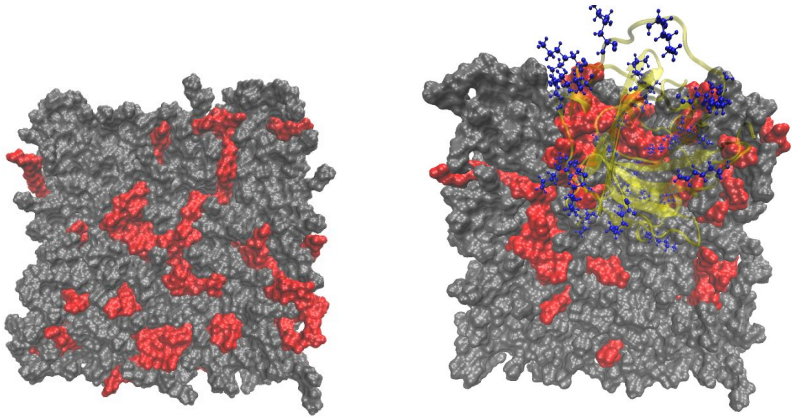


Figure 5.8: Distribution of DLPS on the upper layer in absence (membrane before assembling with protein) and in presence of NP7 (Carmen, after 400 ns). In grey: PC. In red: PS. In yellow: NP7. In blue: Lysines. (VMD)

different selections (Table 5.3) the following remarks can be done:

- the β-barrel is very stable, especially in the open structures;
- a similarity between closed and open structures is apparent;
- the displacement of the heme in the open form is similar but there are differences in the closed structures (see below, Figure 5.17);
- the segment composed by residues 141-154 (see Section 5.1) is conformationally insensitive, its RMSD remains under the average value in every case and in the cross open-closed RMSD it is even lower than the β-barrel's.

5.3.3 Comparison with experimental data

As it was mentioned above (Section 5.1), attention was devoted to residues Arg141, Ala145, Pro147 and Ala154, residues that were NMR-labelled and gave different signals between the in-solution and membrane-bound situations.

For the last three residues, this experimental fact is easily explained in terms of the interaction model imagined, as they all belong to the loop preceding, and to the α-helix singled out as the protein's part mainly involved in the protein-membrane interaction, and would therefore be held compressed between the protein core and the membrane as the interaction takes place, while they would enjoy a relatively larger freedom of movement when the protein is not interacting with membranes.

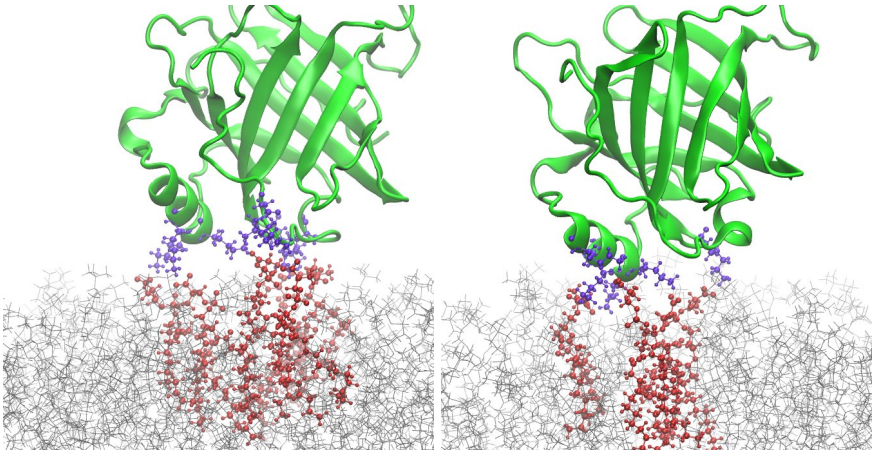


Figure 5.9: Carmen (left) and Azucena (right) docking to the membrane at the end of the SMD phase. (VMD)

In Table 5.4 the differences between the RMSD of these residues and the average protein RMSD are reported. All RMSD are calculated against a single reference structure (final snapshot of *wt*c at 300 ns), counting all heavy atoms (i.e. including sidechains, which is relevant especially for Arg141), using a 100 ns long sampling time (i.e. the last 100 ns of the four trajectories). The difference between the residue value and the overall average gives thus a measure of the residue displacement with respect to the whole protein. For these residues a difference between the *wt*c/o and the Carmen/Azucena values is expected.

A first remark is that in all cases these residues have a lower displacement than the average of the protein (the values are all below zero). Secondly, the membrane-bound values do not always show a lower mobility than the ones in solution, and this suggests that the different behaviour is not simply due to a compression between the membrane and the protein core, as proposed above.

Besides, for the closed structures, data in Table 5.4 show significant differences in Ala145, Pro147 and Ala154, while for the open structures the differences only involve Pro147 and Ala154.

Information about the mobility of this segment of the protein is shown in Figure 5.11, representing the backbone RMSF of the closed and open trajectories both in interaction with the membrane and in solution. While in the closed forms there is a definite difference concerning the single Asn144 and Pro147 residues, in the open structures the whole α-helix, starting with Pro147, displays an overall different mobility, which is reflected in a reduced fluctuation.

Generally speaking for Ala145, Pro147 and Ala154, therefore, the model might be in partial agreement with the NMR data, even though the magnitude of the geometrical changes is only moderate.

	*wt*c	Azucena	*wt*o
Carmen	1.996	3.151	3.260
	1.632	1.274	1.120
	1.524	2.260	2.418
*wt*c		2.858	2.896
		1.537	1.537
		1.964	2.013
Azucena			1.739
			1.055
			0.680

Table 5.3: Cross-structure RMSD (in Å) for membrane-bound and in-solution structures, final snapshots. In black: full backbone atoms. In green: residues 141-154 backbone atoms. In red: β-barrel backbone atoms. (VMD-RMSDTT)

	Carmen	*wt*c	Azucena	*wt*o
Arg141	-1.271	-1.169	-2.360	-2.331
Ala145	-0.297	-0.918	-1.287	-1.229
Pro147	-0.896	-1.218	-2.406	-1.833
Ala154	-1.638	-1.296	-2.031	-2.335

Table 5.4: Difference (in Å) between the heavy-atom RMSD of single NMR-relevant residues (vs. final *wt*c structure, averaged over the last 100 ns) and average heavy-atom RMSD (same reference, same sampling time).

Arg141, on the contrary, does not show the expected differences.

Direct inspection shows that in all cases, Arg141's sidechain is trapped in a network of interactions with Lys19 and Lys116, with Ser22 and Thr24 (Figure 5.12), and this stabilizes its sidechain's position throughout all the trajectories. This explains why the RMSD data show no real difference between in-solution and membrane-bound NP7 as far as Arg141 is concerned.

A suggestion from the NMR data concerned a possible torsion of the C_β-C_γ or C_γ-C_δ bonds upon liposome binding, which would be instrumentally detectable. While the latter bond does not reveal any particular difference between the in-solution and the membrane-bound structures, the distribution plot (Figure 5.13) of the dihedral angle CA-CB-CG-CD, i.e. the angle between the two planes defined by carbon atoms C_α, C_β and C_γ on one side, and C_β, C_γ and C_δ on the other, shows a definite difference between Carmen and *wt*c, and a very similar pattern between *wt*o and Azucena; for the open structures there is also a trend that is inverted with respect to the closed structures (i.e.: in Carmen there is a strong 180° peak not appearing in *wt*c; in *wt*o there is a

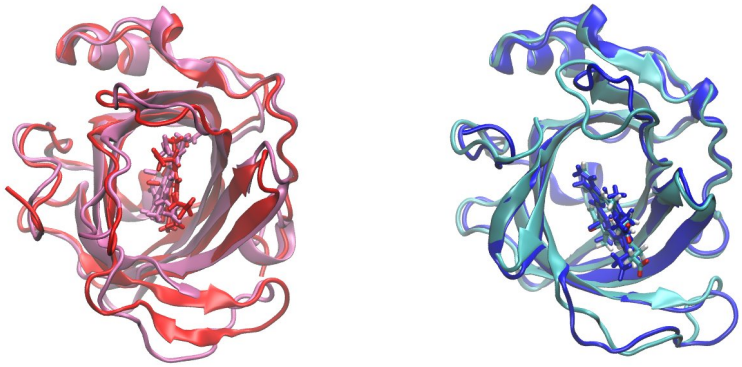


Figure 5.10: Backbone superposition of Carmen (mauve)/*wt*c (red) and Azucena (cyan) /*wt*o (blue). (VMD)

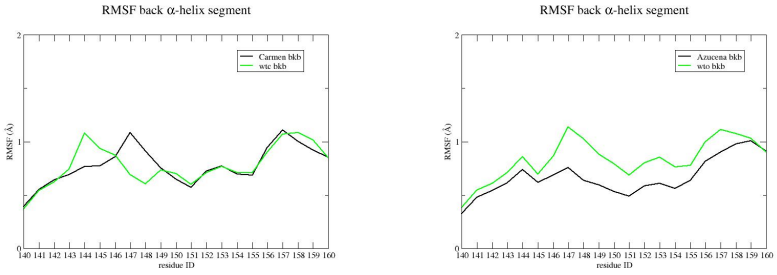


Figure 5.11: Comparison of the RMSF of the back α-helix region. Left: Carmen and *wt*c. Right: Azucena and *wt*o. (VMD)

weak 180° peak not appearing in Azucena).

Naturally this different orientation of the sidechain stem, even if not modifying the mobility of the sidechain itself trapped in multiple hydrogen bonds, can be originated by small structural rearrangements due to the vicinity of the membrane, and including the already mentioned different position of the heme, or changes in the orientation of the hydrogen-bonding residues Ser22 and Thr24, as can be seen in Figure 5.14.

It is possible that the differences in the NMR signals are due to large rearrangement phenomena occurring on longer timescales than those simulated hereby. It seems likely, though, that the presence of an electrically charged membrane so near the rear of the protein can cause detectable shifts in the signals, and that the observed shifts for Arg141 can be explained at least in a large measure, rather than to the loss of the H-bonds or other drastic geomet-

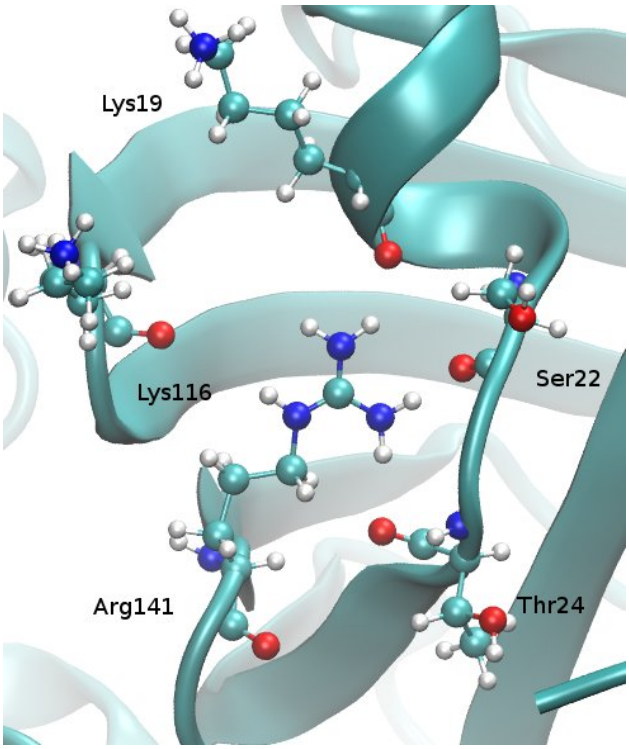


Figure 5.12: Arg141 sidechain hydrogen bond network in Carmen. (VMD)

rical changes, as the effects of either of, possibly both, these causes:

1. change in the hydrogen bond pattern keeping the sidechain blocked
2. torsion in the C_β-C_γ bond due to small rearragement effects.

The NMR data apparently can be justified in the *in silico* model, and this speaks in favour of the quality of the model used. The same model can then be used to try and extend the knowledge of the NP7-membrane system, especially in what concerns the ligand migration paths.

5.3.4 MDpocket, ILS and free ligand MD

In order to obtain the MDpocket results (Table 5.5), two time windows were used for both methods and for both Carmen and Azucena structures. For the MDpocket runs only the protein structure was used, as the membrane does not have a direct steric influence on the protein interior, while for the ILS calculations (Tables 5.6 and 5.7) the system included the membrane as electrostatic potentials are involved in determining the energetic profiles.

The time windows for both Carmen and Azucena structures were the 330-340 ns (s1) and the 370-380 ns (s2) ones.

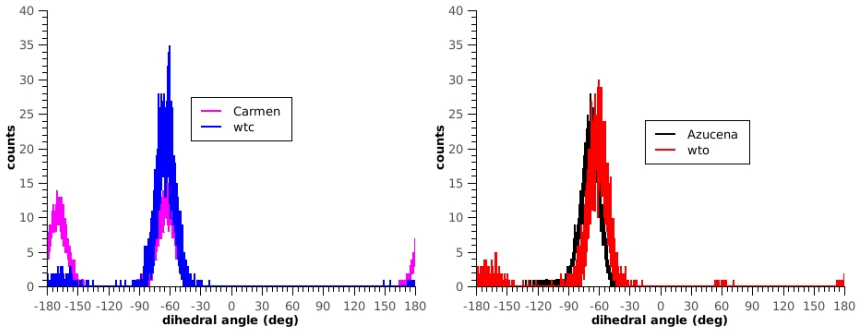


Figure 5.13: CA-CB-CG-CD dihedral angle distribution for Arg141 in Carmen/Azucena and *wtc*/o. (Qtiplot)

	back tunnel	cage[a]-to-back	cage[a]-to-front	cage[a]-to-top
wtc s1	80%	<10%	28%	50%
wtc s2	34%	<10%	26%	55%
Carmen s1	39%	<10%	21%	36%
Carmen s2	44%	<10%	53%	39%
wto	27%	<10%	100%	0%
Azucena s1	18%	<10%	100%	0%
Azucena s2	20%	<10%	100%	0%

Table 5.5: MDpocket frequencies for tunnels in Carmen and Azucena. [a] For "cage" the hydrophobic cage is meant, which is completely disrupted by Asp32 deprotonation's effects in *wto*; in this case the proximity of the CO ligand is considered. See Figure 4.10.

As can be seen, both methods indicate that the connection between the heme pocket and the back of the protein via the axial tunnel is affected by the interaction with membrane.

First, for the open species the connection between sites 1 and 3 seems to be slightly less effective upon binding to the membrane than in solution. Thus, the simulation with free NO indicate that the ligand remains in site 3 or is released to the bulk solvent when it was located in site 1.

Second, in the closed species, the data derived from MDpocket and ILS calculations suggest similar qualitative trends for the protein in the membrane and in solution. However, the simulations performed with free NO point out a distinct behaviour. Thus, whereas in solution there is a passage of the ligand from site 1 to site 3 (and viceversa), the results for the protein attached to the membrane indicate that the ligand has a strong preference for moving from site

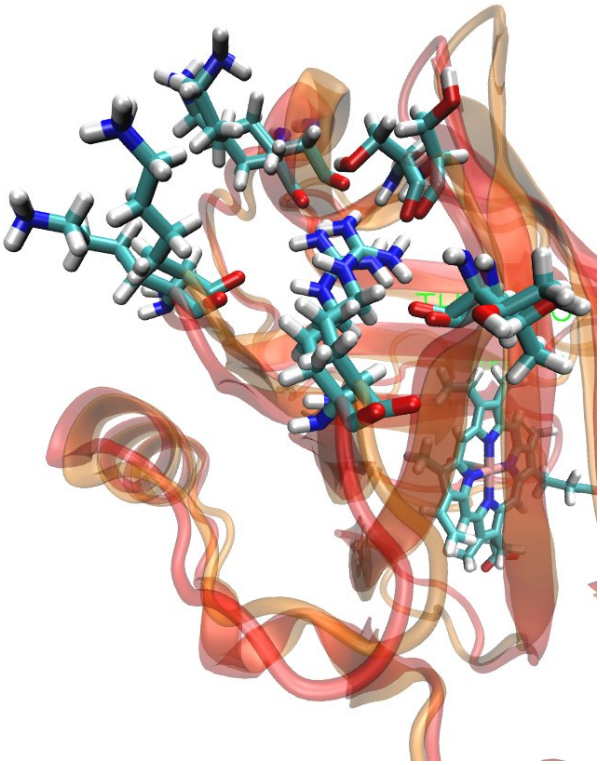


Figure 5.14: Arg141 and its neighbours in Carmen and *wt*c. (VMD)

1 to site 3, while the opposite displacement is not observed.

This seems to indicate that:

- the structural strain that is induced on the protein upon membrane binding has a consequence of a slight "freezing" of the cavity lining sidechains; this is visible comparing the mobility of the residues lining the 1, 2 and 3 cavities in Figure 5.16, where the mobility loss is particularly visible for the residues lining cavities 1 for both open and closed structures, and for those lining cavity 3 in the closed structures, though in this latter case it is accompanied with a moderate increase in the fluctuations in site 2;

- the conformation the sidechains are "frozen"into does not allow the high energetic barriers between inner sites to temporarily drop;

- the energetic barrier between the distal pocket and site 1 is greatly increased because of a deeper and slightly rotated position of the heme in the barrel in Carmen than in *wt*c (Figure 5.17) which allows less space between the pyrrole-I (see Figure 1.3) vinyl and the neighbouring sidechains.

	1a	1b	1c	2	3	4
*wt*c s1	-3.9	+2.4	+14	-5.8	-4.7	+8.0
*wt*c s2	-1.9	+3.0	+14	-4.2	-5.9	+7.1
Carmen s1	-3.9	+0.1	+2.4	-3.2	-3.8	-4.2
Carmen s2	-4.9	+0.7	-1.2	-2.0	-4.8	-4.8
*wt*o	-6.4	-1.8	-3.2	-1.0	-5.6	+0.3
Azucena s1	-5.7	-0.5	-4.4	-1.1	-5.9	+0.6
Azucena s2	-4.9	-1.9	-1.7	-1.7	-5.2	+0.0

Table 5.6: ILS isovalues for the docking sites of *wt*NP7 in solution and membrane-interacting.

	BT	SED	SEA	CTB	BTS	BTF
*wt*c s1	+2.6	+10[a]	+19	+60	+60	+29
*wt*c s2	+5.3	+5.4[a]	+18	+50	+53	+30
Carmen s1	+3.4	+9.8[a]	+3.4	>85	+85	+19
Carmen s2	+3.2	+22[a]	-1.0	>85	+85	+15
*wt*o	+10	-	+11	+10	+10	+50
Azucena s1	+23	-	+14	+23	+23	+44
Azucena s2	+21	-	+14	+21	+21	+65

Table 5.7: ILS isovalues for the tunnels of *wt*NP7 in solution and membrane-interacting. [a] "Top" connected at lower isovalues. For the tunnels definitions, see 4.3.2.

On the other hand, while in Azucena the direct exposure of the heme to the solvent is already high being similar to that of *wt*o, in Carmen the in-cage energetic minimum (site 4) is much deeper than in NP7 in solution, and the SEA (frontal) escape route is significantly favoured.

5.4 Conclusions

Membrane-binding is a unique capability exhibited by NP7 alone among nitrophorins.

This study could prove that indeed the Lys higher concentration at the back of the barrel is related to such capability. Furthermore, it could show the important role of the α-helix segment in the protein-membrane interaction process.

Seeing the impact the interaction with the membrane has upon the axial tunnel, a suggestion can be made about a possible functional distinction of NP7 in the context of the Rp nitrophorins. It is not clear yet how membrane binding can influence the pH value at which the conformational transition (related,

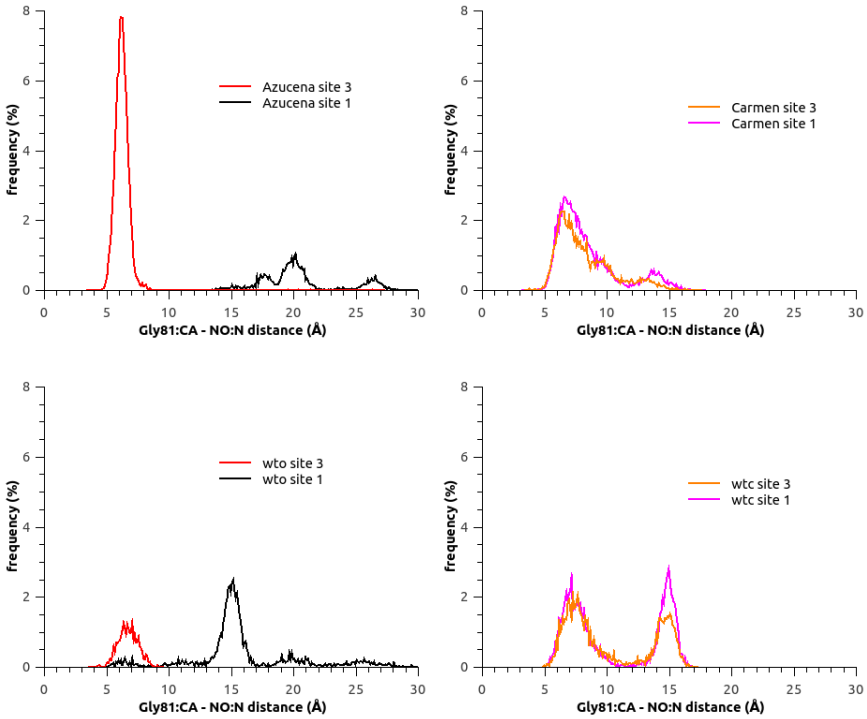


Figure 5.15: Free NO distributions for membrane-bound NP7. Top left: open structure (Azucena), NO starting position site 1 (black) and site 3 (red). Top right: closed structure (Carmen), NO starting position site 1 (magenta) and site 3 (orange). Bottom: reference profiles for *wt*o (left) and *wt*c (right). In all plots the counts were normalized by the number of planned MD steps, in order to also visualize the ligand's residence inside the axial tunnel or its escape. The plots on *wt*NP7 summarize the data shown in Figure 3.6. (Qtiplot)

as discussed in Chapter 3, to the protonation state of Asp32, its consequent interaction with Ile132 and the positioning of loops AB and GH) or the NO affinity, and whether it may be triggering NO release or not. Nevertheless, in a membrane-bound NP7 the communication between the heme pocket and the interior of the protein is made more difficult, while the frontal exchange with the solvent easier. All this seems to indicate the possibility that, if nitrophorins are a family of proteins releasing vasodilating NO on different timescales into the blood vessels during Rp's meal, NP7 might be the one whose time frame is the last, i.e. when the immune system has detected the presence of endothelial damage and activated platelets begin repairing the wound through which the insect is pumping saliva in and blood out.

In other words, if the picture offered by the present study is correct, and this can only be determined through deeper and more specific experimental

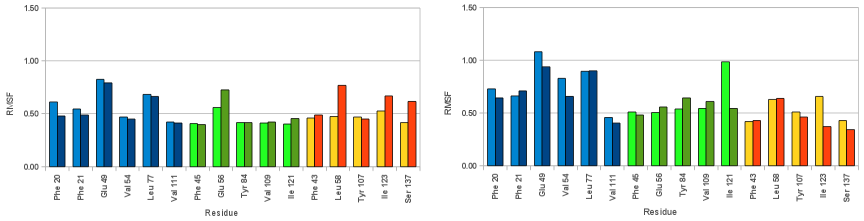


Figure 5.16: Comparison of RMSF (Å) values for the open structures (left: bright colours for *wt*o, dark colours for Azucena) and for the closed structures (right: bright colours for *wt*c, dark colours for Carmen). (OpenOffice)

research on membrane-bound NP7, the protein's function might be to release NO at the last stage of the meal, to keep a sufficient blood flow even while the bite is starting to be healed by the victim's defensive devices. There are other possible alternative or additional interpretations, such as selective and specific binding to membrane proteins, or interference with platelet aggregation and clumping through shielding and concentration of negative charges.

This functionality hypothesis is consistent with the fact that NP7 is only expressed in nearly-adult or adult insects, i.e. when the feeding needs are higher and the chances of being detected and hunted by the victim during a longer meal are also higher. The evolutionary strategy of deploying different nitrophorins might then be seen as aimed at keeping the NO release constant during all the time of the meal, in order to increase the blood flow, block the intrusion messengers as histamine and the biochemical cascades leading to blood coagulation and wound repair. In this picture, NP7 might as well be seen as the last one to release NO: when coagulation has started, and the insect's meal is at its end, NP7 allows Rp a dessert.

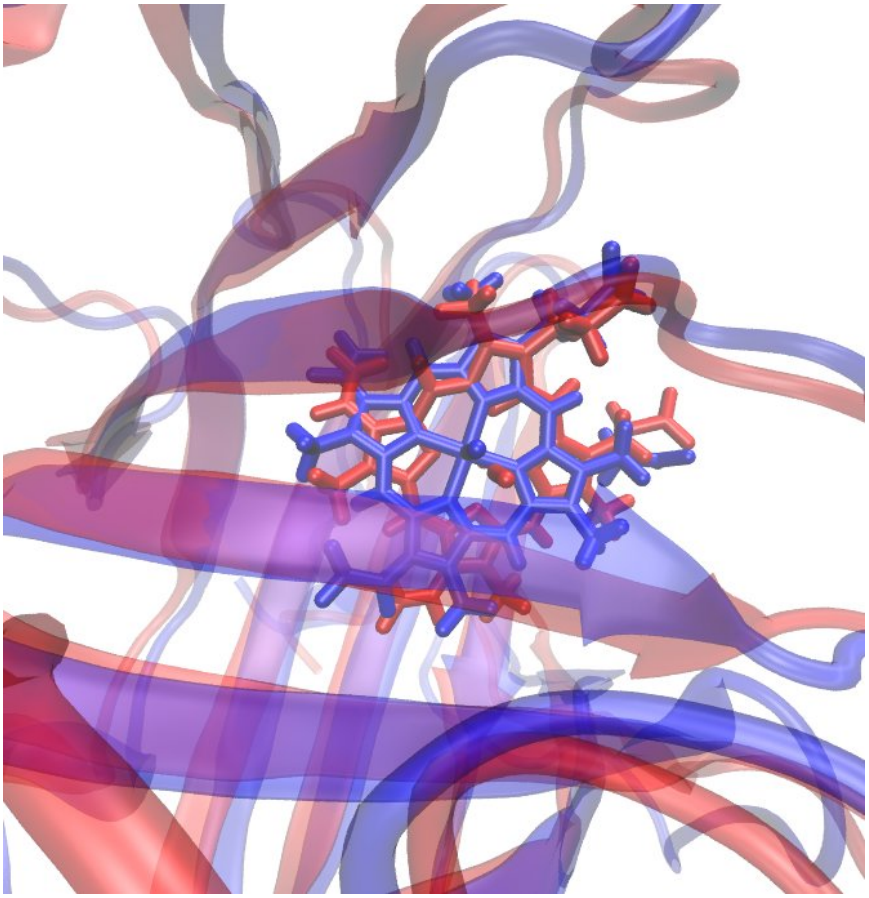


Figure 5.17: Comparison between the heme's position in Carmen (blue) and in *wt*c (red). (VMD)

Acknowledgments and thanks

The Instituto de Astrofísica de Canarias (IAC), the Barcelona Supercomputer Center (BSC) and the Centre de Serveis Científics i Acadèmics de Catalunya (CESCA) are aknowledged for providing the computational resources for the present studies.

The present research work was partially funded by:

- Progetto azioni integrate Italia-Spagna 2009 (MIUR/Ministerio de Educación y Ciencia): "Migrazione di ligandi attraverso le cavità idrofobiche di emoglobine non simbiotiche di Arabidopsis thaliana";
- Programma Vigoni 2009 (MIUR/Ateneo Italo-Tedesco/DAAD): "Mechanism of the Release of Nitric Oxide from the Cell Surface Binding Ferriheme Protein Nitrophorin 7".

During the past three years I have had plenty of reasons to be grateful, and several people to thank. Now, as this period of fascinating and engaging research work is over and I return to my high school teaching position, I can only mention few of the exceptional people I have had the opportunity to meet and work with. My gratitude to them goes beyond the professional realm, and shall therefore remain on this page largely unexpressed; this holds especially true for the leader of the research group of which I have been honoured to be a member, and for my tutors.

Among the many things for which I can thank Dr. Stefania Abbruzzetti I would mention here her sometimes surprising, always motivating and uplifting confidence in my work and her commitment to promote it.

My heartfelt thanks go to Prof. F. Javier Luque for being a most welcoming host in Barcelona, and for all I learnt from him, as a mentor, but also as an example of constant care, dedication and support even when it came to a PhD student with little experience and only occasionally visiting from abroad. Being able to work with him and his great group really was a privilege.

Many thanks go to Dr. Axel Bidon-Chanal (Universitat de Barcelona): his thorough knowledge of the technical aspects of the computational jobs, his

keen eye and deep insight in the subject topics and his helpfulness saved my day on many occasions.

My gratitude also flies to Germany, to Dr. Markus Knipp (Max Planck Institut für chemische Energiekonversion, Mülheim an der Ruhr), whose interest and expertise in NPs and in particular in NP7 were crucial in several cases; and to Prof. Henrike Heise (Heinrich-Heine Universität Düsseldorf and Forschungszentrum Jülich) for her patience and clarity in providing and explaining the NMR signals left by our protein.

I am grateful to Dr. Francesca Spyrakis (now at Università di Modena e Reggio Emilia) who introduced me to the charming world of Molecular Dynamics, for all the support and information she provided, including her admiration for Barcelona's group as bonus material.

Many thanks to Dr. Stefano Bruno (Dipartimento di Farmacia, Università di Parma) for taking care of the laboratory issues posed by the preparation of our reluctant NP7, looking forward to future occasions of discussing again of tricks-of-the-trade or travels while enjoying a weizen.

Prof. Giulio Cerullo and particularly Dr. Giancarlo Soavi (Politecnico di Milano) have been very supportive and the full availability they gave for ultrafast rebinding kinetics measurements is greatly appreciated.

Thanks also to Dr. Ana Oliveira for stimulating discussions, to Dr. Juan Pablo Bustamante for sharing ILS tips and tricks; many thanks to Dr. Carmen Mandalari, because walking on parallel lines perhaps means never crossing each other's path, but it also means that you can count on someone else nearby for a chat, an opinion, a piece of advice, or just a coffee.

During these three years two members of my family left me. Prof. Giancarla Allegri was the one who first made me enter a laboratory when I was a child of four or five, allowed me to look into a microscope, showed me slides and coltures, introduced me to chicks (in the literal sense). On my aunt's face I first saw love for science, passion for research, respect for the University institutions.

Among all the memories I have of the second person, there is a faint and distant one of him laying books on the dining-room table late in the evening, so that he could study for his degree in his quiet moment of the day, after work, and after kissing me, his firstborn, goodnight. He could only see the beginning of my doctorate, but his example has been guiding me despite his departure.

I want to mention both of them at the end of this work, because their inspiration helped me through it.

I feel I have with my colleague prof. Raffaella Dall'Aglio a debt of gratitude, as she was the one who first suggested to me the possibility of a leave of absence for attending a PhD programme, and encouraged me to pursue this option. At the same time I would also like to express my regret for the restric-

tive amendments brought in the meantime to the law allowing this, as I am absolutely sure that a teacher's professional skills, methods, credibility and authoritativeness have much to gain by him or her devoting time to studying and researching.

Thanks to my mother's help, in terms of interest and logistics, a few things hardly likely to happen became possible (among which I shall mention my visits to Spain and Germany) and many difficult ones easier.

Finally, my wholehearted thanks to Hilary, my wife and companion: not only has her loving, unfailing, substantial support on a project we shared since its very beginning, made arriving at these last words possible, but it also has shaped and wrapped it as a gift to our family.

Bibliography

[1] P.A. Buxton. The biology of a blood-sucking bug, *Rhodnius Prolixus* Stahl. *Trans. R. Soc. Entomol. Lond.*, 78:227–236, 1930.

[2] L.V. Kirchhoff. American trypanosomiasis (Chagas' disease)–a tropical disease now in the United States. *N. Engl. J. Med.*, 329(9):639–44, August 1993. ISSN 0028-4793. doi: 10.1056/NEJM199308263290909.

[3] J.M. Ribeiro and E.S. Garcia. Platelet antiaggregating activity in the salivary secretion of the blood sucking bug *Rhodnius prolixus*. *Experientia*, 37(4):384–6, April 1981. ISSN 0014-4754. URL `http://www.ncbi.nlm.nih.gov/pubmed/7016577`.

[4] J.M. Ribeiro. The antiserotonin and antihistamine activities of salivary secretion of *Rhodnius prolixus*. *J. Insect Physiol.*, 28(1):69–75, 1982. ISSN 0022-1910. doi: 10.1016/0022-1910(82)90024-5.

[5] J.M. Ribeiro, J.M. Hazzard, R.H. Nussenzveig, D.E. Champagne, and F.A. Walker. Reversible binding of nitric oxide by a salivary heme protein from a bloodsucking insect. *Science*, 260(5107):539–41, April 1993. ISSN 0036-8075. doi: 10.1126/science.8386393.

[6] J.M. Ribeiro and F.A. Walker. High affinity histamine-binding and antihistaminic activity of the salivary nitric oxide-carrying heme protein Nitrophorin of *Rhodnius prolixus*. *J. Exp. Med.*, 180(6):2251–7, December 1994. doi: 10.1084/jem.180.6.2251.

[7] F.A. Walker. Nitric oxide interaction with insect Nitrophorins and thoughts on the electron configuration of the $\{FeNO\}^6$ complex. *J. Inorg. Biochem.*, 99(1):216–36, January 2005. ISSN 0162-0134. doi: 10.1016/j.jinorgbio.2004.10.009.

[8] D.E. Champagne, R.H. Nussenzveig, and J.M. Ribeiro. Purification, partial characterization, and cloning of nitric oxide-carrying heme proteins (Nitrophorins) from salivary glands of the blood-sucking insect *Rhodnius prolixus*. *J. Biol. Chem.*, 270(15):8691–5, April 1995. doi: 10.1074/jbc.270.15.8691.

[9] K. Hellmann and R.I. Hawkins. Prolixin-S and Prolixin-G; two anticoagulants from *Rhodnius prolixus Stål*. *Nature*, 207(994):265–7, July 1965. ISSN 0028-0836. doi: 10.1038/207265a0.

[10] J.M. Ribeiro, M. Schneider, and J.A. Guimarães. Purification and characterization of Prolixin S (Nitrophorin 2), the salivary anticoagulant of the blood-sucking bug *Rhodnius prolixus*. *Biochem. J.*, 308 (Pt 1):243–9, May 1995. ISSN 0264-6021. URL `http://www.ncbi.nlm.nih.gov/pubmed/7755571`.

[11] J. Sun, M. Yamaguchi, M. Yuda, K. Miura, H. Takeya, M. Hirai, H. Matsuoka, K. Ando, T. Watanabe, K. Suzuki, and Y. Chinzei. Purification, characterization and cDNA cloning of a novel anticoagulant of the intrinsic pathway, (Prolixin-S) from salivary glands of the blood sucking bug, *Rhodnius prolixus*. *Thromb. Haemost.*, 75(4):573–7, April 1996. ISSN 0340-6245. URL `http://www.ncbi.nlm.nih.gov/pubmed/8743181`.

[12] Y. Zhang, J.M. Ribeiro, J.A. Guimarães, and P.N. Walsh. Nitrophorin-2: a novel mixed-type reversible specific inhibitor of the intrinsic factor-X activating complex. *Biochemistry*, 37(30):10681–90, July 1998. ISSN 0006-2960. doi: 10.1021/bi973050y.

[13] Y. Kaneko, M. Yuda, T. Iio, T. Murase, and Y. Chinzei. Kinetic analysis on nitric oxide binding of recombinant Prolixin-S, a nitric oxide transport protein from the bloodsucking bug, *Rhodnius prolixus*. *Biochim. Biophys. Acta*, 1431(2):492–9, May 1999. ISSN 0006-3002. doi: 10.1016/S0167-4838(99)00082-5.

[14] J. Sun, M. Yuda, K. Miura, and Y. Chinzei. Characterization and cDNA cloning of a hemoprotein in the salivary glands of the blood-sucking insect, *Rhodnius prolixus*. *Insect Biochem. Mol. Biol.*, 28(3):191–200, March 1998. ISSN 0965-1748. doi: 10.1016/S0965-1748(97)00115-X.

[15] M. Yuda, K. Higuchi, J. Sun, Y. Kureishi, M. Ito, and Y. Chinzei. Expression, reconstitution and characterization of Prolixin-S as a vasodilator – a salivary gland nitric-oxide-binding hemoprotein of *Rhodnius prolixus*. *Eur. J. Biochem.*, 249(1):337–42, October 1997. ISSN 0014-2956. doi: 10.1111/j.1432-1033.1997.00337.x.

[16] J.F. Andersen, D.E. Champagne, A. Weichsel, J.M. Ribeiro, C.A. Balfour, V. Dress, and W.R. Montfort. Nitric oxide binding and crystallization of recombinant Nitrophorin I, a nitric oxide transport protein from the blood-sucking bug *Rhodnius prolixus*. *Biochemistry*, 36(15):4423–8, April 1997. doi: 10.1021/bi9628883.

[17] A. Weichsel, J.F. Andersen, D.E. Champagne, F.A. Walker, and W.R. Montfort. Crystal structures of a nitric oxide transport protein from a blood-sucking insect. *Nat. Struct. Biol.*, 5(4):304–9, April 1998. doi: 10.1038/nsb0498-304.

[18] X.D. Ding, A. Weichsel, J.F. Andersen, T.K. Shokhireva, C.A. Balfour, A.J. Pierik, B.A. Averill, W.R. Montfort, and F.A. Walker. Nitric oxide binding to the ferri- and ferroheme states of Nitrophorin 1, a reversible NO-binding heme protein from the saliva of the blood-sucking insect, *Rhodnius prolixus*. *J. Am. Chem. Soc.*, 121(1):128–138, 1999. doi: 10.1021/ja982979i.

[19] J.F. Andersen, A. Weichsel, C.A. Balfour, D.E. Champagne, and W.R. Montfort. The crystal structure of Nitrophorin 4 at 1 5Å resolution: transport of nitric oxide by a Lipocalin-based heme protein. *Structure*, 6 (10):1315–27, October 1998. doi: 10.1016/S0969-2126(98)00131-2.

[20] J.F. Andersen, X.D. Ding, C.A. Balfour, T.K. Shokhireva, D.E. Champagne, F.A. Walker, and W.R. Montfort. Kinetics and equilibria in ligand binding by Nitrophorins 1-4: evidence for stabilization of a nitric oxide-ferriheme complex through a ligand-induced conformational trap. *Biochemistry*, 39(33):10118–31, August 2000. doi: 10.1021/bi000766b.

[21] A. Weichsel, J.F. Andersen, S.A. Roberts, and W.R. Montfort. Nitric oxide binding to Nitrophorin 4 induces complete distal pocket burial. *Nat. Struct. Biol.*, 7(7):551–4, July 2000. ISSN 1072-8368. doi: 10.1038/76769.

[22] J.F. Andersen and W.R. Montfort. The crystal structure of Nitrophorin 2. A trifunctional antihemostatic protein from the saliva of *Rhodnius prolixus*. *J. Biol. Chem.*, 275(39):30496–503, September 2000. doi: 10.1074/jbc.M002857200.

[23] W.R. Montfort, A. Weichsel, and J.F. Andersen. Nitrophorins and related antihemostatic Lipocalins from *Rhodnius prolixus* and other blood-sucking arthropods. *Biochim. Biophys. Acta*, 1482(1-2):110–8, October 2000. doi: 10.1016/S0167-4838(00)00165-5.

[24] I.M.B. Francischetti, J.M. Ribeiro, D.E. Champagne, and J.F. Andersen. Purification, cloning, expression, and mechanism of action of a novel platelet aggregation inhibitor from the salivary gland of the blood-sucking bug, *Rhodnius prolixus*. *J. Biol. Chem.*, 275(17):12639–50, April 2000. ISSN 0021-9258. doi: 10.1074/jbc.275.17.12639.

[25] I.M.B. Francischetti, J.F. Andersen, and J.M. Ribeiro. Biochemical and functional characterization of recombinant *Rhodnius prolixus* platelet

aggregation inhibitor 1 as a novel Lipocalin with high affinity for adenosine diphosphate and other adenine nucleotides. *Biochemistry*, 41(11): 3810–8, March 2002. ISSN 0006-2960. doi: 10.1021/bi011015s.

[26] H. Isawa, M. Yuda, K. Yoneda, and Y. Chinzei. The insect salivary protein, Prolixin-S, inhibits factor IXa generation and Xase complex formation in the blood coagulation pathway. *J. Biol. Chem.*, 275(9):6636–41, March 2000. ISSN 0021-9258. doi: 10.1074/jbc.275.9.6636.

[27] S.A. Roberts, A. Weichsel, Y. Qiu, J.A. Shelnutt, F.A. Walker, and W.R. Montfort. Ligand-induced heme ruffling and bent no geometry in ultra-high-resolution structures of Nitrophorin 4. *Biochemistry*, 40(38): 11327–37, September 2001. doi: 10.1021/bi0109257.

[28] T.K. Shokhireva, R.E. Berry, E. Uno, C.A. Balfour, H. Zhang, and F.A. Walker. Electrochemical and NMR spectroscopic studies of distal pocket mutants of Nitrophorin 2: stability, structure, and dynamics of axial ligand complexes. *Proc. Natl. Acad. Sci. U.S.A.*, 100(7):3778–83, April 2003. doi: 10.1073/pnas.0536641100.

[29] E.M. Maes, F.A. Walker, W.R. Montfort, and R.S. Czernuszewicz. Resonance Raman spectroscopic study of Nitrophorin 1, a nitric oxide-binding heme protein from *Rhodnius prolixus*, and its nitrosyl and cyano adducts. *J. Am. Chem. Soc.*, 123(47):11664–72, November 2001. ISSN 0002-7863. doi: 10.1021/ja0031927.

[30] J.F. Andersen, I.M.B. Francischetti, J.G. Valenzuela, P. Schuck, and J.M. Ribeiro. Inhibition of hemostasis by a high affinity biogenic amine-binding protein from the saliva of a blood-feeding insect. *J. Biol. Chem.*, 278(7):4611–7, February 2003. doi: 10.1074/jbc.M211438200.

[31] J.M. Ribeiro, J.F. Andersen, M.A.C. Silva-Neto, V.M. Pham, M.K. Garfield, and J.G. Valenzuela. Exploring the sialome of the blood-sucking bug *Rhodnius prolixus*. *Insect Biochem. Mol. Biol.*, 34(1):61–79, January 2004. ISSN 0965-1748. doi: 10.1016/j.ibmb.2003.09.004.

[32] M.F. Moreira, H.S. Coelho, R.B. Zingali, P.L. Oliveira, and H. Masuda. Changes in salivary Nitrophorin profile during the life cycle of the blood-sucking bug *Rhodnius prolixus*. *Insect Biochem. Mol. Biol.*, 33(1):23–8, January 2003. ISSN 0965-1748. doi: 10.1016/S0965-1748(02)00163-7.

[33] J.F. Andersen, N.P. Gudderra, I.M.B. Francischetti, J.G. Valenzuela, and J.M. Ribeiro. Recognition of anionic phospholipid membranes by an antihemostatic protein from a blood-feeding insect. *Biochemistry*, 43 (22):6987–94, June 2004. doi: 10.1021/bi049655t.

[34] P. Ascenzi, M. Nardini, M. Bolognesi, and W.R. Montfort. Nitrophorins: Lipocalin-based heme proteins transporting nitric oxide. *Biochem. Mol. Biol. Educ.*, 30:68–71, 2002. doi: 10.1002/bmb.2002.494030010016.

[35] E.M. Maes, A. Weichsel, J.F. Andersen, D. Shepley, and W.R. Montfort. Role of binding site loops in controlling nitric oxide release: structure and kinetics of mutant forms of Nitrophorin 4. *Biochemistry*, 43(21): 6679–90, June 2004. doi: 10.1021/bi049748a.

[36] K. Nienhaus, E.M. Maes, A. Weichsel, W.R. Montfort, and G.U. Nienhaus. Structural dynamics controls nitric oxide affinity in Nitrophorin 4. *J. Biol. Chem.*, 279(38):39401–7, September 2004. doi: 10.1074/jbc.M406178200.

[37] D.A. Kondrashov, S.A. Roberts, A. Weichsel, and W.R. Montfort. Protein functional cycle viewed at atomic resolution: conformational change and mobility in Nitrophorin 4 as a function of pH and NO binding. *Biochemistry*, 43(43):13637–47, November 2004. doi: 10.1021/bi0483155.

[38] A. Benabbas, X. Ye, M. Kubo, Z. Zhang, E.M. Maes, W.R. Montfort, and P.M. Champion. Ultrafast dynamics of diatomic ligand binding to Nitrophorin 4. *J. Am. Chem. Soc.*, 132(8):2811–20, feb 2010. doi: 10.1021/ja910005b.

[39] R.E. Berry, X.D. Ding, T.K. Shokhireva, A. Weichsel, W.R. Montfort, and F.A. Walker. Axial ligand complexes of the *Rhodnius* Nitrophorins: reduction potentials, binding constants, EPR spectra, and structures of the 4-iodopyrazole and imidazole complexes of NP4. *J. Biol. Inorg. Chem.*, 9(2):135–44, March 2004. ISSN 0949-8257. doi: 10.1007/s00775-003-0505-0.

[40] D.K. Menyhárd and G.M. Keserü. Protonation state of Asp30 exerts crucial influence over surface loop rearrangements responsible for NO release in Nitrophorin 4. *FEBS Lett.*, 579(24):5392–8, October 2005. doi: 10.1016/j.febslet.2005.09.003.

[41] E.M. Maes, S.A. Roberts, A. Weichsel, and W.R. Montfort. Ultrahigh resolution structures of Nitrophorin 4: heme distortion in ferrous CO and NO complexes. *Biochemistry*, 44(38):12690–9, September 2005. ISSN 0006-2960. doi: 10.1021/bi0506573.

[42] J.F. Andersen, N.P. Gudderra, I.M.B. Francischetti, and J.M. Ribeiro. The role of salivary lipocalins in blood feeding by *Rhodnius prolixus*. *Arch. Insect. Biochem. Physiol.*, 58(2):97–105, February 2005. ISSN 0739-4462. doi: 10.1002/arch.20032.

[43] A. Ambrus, K. Friedrich, and A. Somogyi. Oligomerization of nitrophorins. *Anal. Biochem.*, 352(2):286–95, May 2006. ISSN 0003-2697. doi: 10.1016/j.ab.2006.02.003.

[44] R.E. Berry, T.K. Shokhireva, I. Filippov, M.N. Shokhirev, H. Zhang, and F.A. Walker. Effect of the N-terminus on heme cavity structure, ligand equilibrium, rate constants, and reduction potentials of Nitrophorin 2 from *Rhodnius prolixus*. *Biochemistry*, 46(23):6830–43, June 2007. ISSN 0006-2960. doi: 10.1021/bi7002263.

[45] T.K. Shokhireva, N.V. Shokhirev, and F.A. Walker. Assignment of heme resonances and determination of the electronic structures of high- and low-spin Nitrophorin 2 by ^{1}H and ^{13}C NMR spectroscopy: an explanation of the order of heme methyl resonances in high-spin ferriheme proteins. *Biochemistry*, 42(3):679–93, January 2003. ISSN 0006-2960. doi: 10.1021/bi026765w.

[46] T.K. Shokhireva, A. Weichsel, K.M. Smith, R.E. Berry, N.V. Shokhirev, C.A. Balfour, H. Zhang, W.R. Montfort, and F.A. Walker. Assignment of the ferriheme resonances of the low-spin complexes of Nitrophorins 1 and 4 by ^{1}H and ^{13}C NMR spectroscopy: comparison to structural data obtained from X-ray crystallography. *Inorg. Chem.*, 46(6):2041–56, March 2007. doi: 10.1021/ic061408l.

[47] T.K. Shokhireva, K.M. Smith, R.E. Berry, N.V. Shokhirev, C.A. Balfour, H. Zhang, and F.A. Walker. Assignment of the ferriheme resonances of the high-spin forms of Nitrophorins 1 and 4 by ^{1}H NMR spectroscopy: comparison to structural data obtained from X-ray crystallography. *Inorg. Chem.*, 46(1):170–8, January 2007. ISSN 0020-1669. doi: 10.1021/ic061407t.

[48] T.K. Shokhireva, R.E. Berry, H. Zhang, N.V. Shokhirev, and F.A. Walker. Assignment of ferriheme resonances for high- and low-spin forms of Nitrophorin 3 by H and C NMR spectroscopy and comparison to Nitrophorin 2: heme pocket structural similarities and differences. *Inorg. Chimica Acta*, 361(4):925–940, March 2008. ISSN 0020-1693. doi: 10.1016/j.ica.2007.05.044.

[49] A.M. Amoia and W.R. Montfort. Apo-nitrophorin 4 at atomic resolution. *Protein Sci.*, 16(9):2076–81, September 2007. ISSN 0961-8368. doi: 10.1110/ps.072981907.

[50] D.A. Kondrashov and W.R. Montfort. Nonequilibrium dynamics simulations of nitric oxide release: comparative study of Nitrophorin and Myoglobin. *J. Phys. Chem. B*, 111(31):9244–52, August 2007. doi: 10.1021/jp071136n.

[51] M. Meuwly, O.M. Becker, R. Stote, and M. Karplus. NO rebinding to Myoglobin: a reactive molecular dynamics study. *Biophys. Chem.*, 98 (1-2):183–207, July 2002. doi: 10.1016/S0301-4622(02)00093-5.

[52] M.A. Martí, M.C. González-Lebrero, A.E. Roitberg, and D.A. Estrin. Bond or cage effect: how Nitrophorins transport and release nitric oxide. *J. Am. Chem. Soc.*, 130(5):1611–8, February 2008. doi: 10.1021/ja075565a.

[53] R.E. Berry, M.N. Shokhirev, A.Y.W. Ho, F. Yang, T.K. Shokhireva, H. Zhang, A. Weichsel, W.R. Montfort, and F.A. Walker. Effect of mutation of carboxyl side-chain amino acids near the heme on the midpoint potentials and ligand binding constants of Nitrophorin 2 and its NO, histamine, and imidazole complexes. *J. Am. Chem. Soc.*, 131(6):2313–27, February 2009. doi: 10.1021/ja808105d.

[54] M. Kubo, F. Gruia, A. Benabbas, A. Barabanschikov, W.R. Montfort, E.M. Maes, and P.M. Champion. Low-frequency mode activity of heme: femtosecond coherence spectroscopy of iron porphine halides and Nitrophorin. *J. Am. Chem. Soc.*, 130(30):9800–11, July 2008. doi: 10.1021/ja800916d.

[55] F. Yang, H. Zhang, and M. Knipp. A one-residue switch reverses the orientation of a heme b cofactor. Investigations of the ferriheme NO transporters Nitrophorin 2 and 7 from the blood-feeding insect *Rhodnius prolixus*. *Biochemistry*, 48(2):235–41, January 2009. doi: 10.1021/bi8020229.

[56] J.M. Swails, Y. Meng, F.A. Walker, M.A. Martí, D.A. Estrin, and A.E. Roitberg. pH-dependent mechanism of nitric oxide release in Nitrophorins 2 and 4. *J. Phys. Chem. B*, 113(4):1192–201, January 2009. doi: 10.1021/jp806906x.

[57] M. Knipp, H. Zhang, R.E. Berry, and F.A. Walker. Overexpression in *Escherichia coli* and functional reconstitution of the liposome binding ferriheme protein Nitrophorin 7 from the bloodsucking bug *Rhodnius prolixus*. *Protein Expr. Purif.*, 54(1):183–91, July 2007. doi: 10.1016/j.pep.2007.02.017.

[58] M. Knipp, F. Yang, R.E. Berry, H. Zhang, M.N. Shokhirev, and F.A. Walker. Spectroscopic and functional characterization of Nitrophorin 7 from the blood-feeding insect *Rhodnius prolixus* reveals an important role of its isoform-specific N-terminus for proper protein function. *Biochemistry*, 46(46):13254–68, November 2007. doi: 10.1021/bi7014986.

[59] F. Yang. *Nuclear Magnetic Resonance studies on Nitrophorins and paramagnetic model complexes*. PhD thesis, The University of Arizona, 2010.

[60] C. He and M. Knipp. Formation of nitric oxide from nitrite by the ferriheme b protein Nitrophorin 7. *J. Am. Chem. Soc.*, 131(34):12042–3, September 2009. doi: 10.1021/ja9040362.

[61] C. He, H. Ogata, and M. Knipp. Formation of the complex of nitrite with the ferriheme b β-barrel proteins Nitrophorin 4 and Nitrophorin 7. *Biochemistry*, 49(28):5841–51, July 2010. doi: 10.1021/bi100324z.

[62] R.E. Berry, D. Muthu, T.K. Shokhireva, S.A. Garrett, H. Zhang, and F.A. Walker. Native N-terminus Nitrophorin 2 from the kissing bug: similarities to and differences from NP2(D1A). *Chem. Biodivers.*, 9(9):1739–55, September 2012. ISSN 1612-1880. doi: 10.1002/cbdv.201100449.

[63] M. Knipp and C. He. Nitrophorins: nitrite disproportionation reaction and other novel functionalities of insect heme-based nitric oxide transport proteins. *IUBMB Life*, 63(5):304–12, May 2011. doi: 10.1002/iub.451.

[64] M. Knipp, J.J. Taing, and C. He. Reduction of the Lipocalin type heme containing protein Nitrophorin – sensitivity of the fold-stabilizing cysteine disulfides toward routine heme-iron reduction. *J. Inorg. Biochem.*, 105(11):1405–12, November 2011. doi: 10.1016/j.jinorgbio.2011.07.009.

[65] C. He, S. Neya, and M. Knipp. Breaking the proximal Fe(II)-N(His) bond in heme proteins through local structural tension: lessons from the heme b proteins Nitrophorin 4, Nitrophorin 7, and related site-directed mutant proteins. *Biochemistry*, 50(40):8559–75, October 2011. doi: 10.1021/bi201073t.

[66] S. Abbruzzetti, C. He, H. Ogata, S. Bruno, C. Viappiani, and M. Knipp. Heterogeneous kinetics of the carbon monoxide association and dissociation reaction to Nitrophorin 4 and 7 coincide with structural heterogeneity of the gate-loop. *J. Am. Chem. Soc.*, 134(24):9986–98, June 2012. doi: 10.1021/ja2121662.

[67] N.V. Di Russo, D.A. Estrin, M.A. Martí, and A.E. Roitberg. pH-Dependent conformational changes in proteins and their effect on experimental pK_as: the case of Nitrophorin 4. *PLoS Comput. Biol.*, 8(11):e1002761, 2012. doi: 10.1371/journal.pcbi.1002761.

[68] H. Ogata and M. Knipp. Crystallization and preliminary X-ray crystallographic analysis of the membrane-binding haemprotein Nitrophorin 7 from *Rhodnius prolixus*. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.*, 68(Pt 1):37–40, January 2012. ISSN 1744-3091. doi: 10.1107/S1744309111044708.

[69] M. Knipp, R.P. Soares, and M.H. Pereira. Identification of the native N-terminus of the membrane attaching ferriheme protein Nitrophorin 7 from *Rhodnius prolixus*. *Anal. Biochem.*, 424(1):79–81, May 2012. doi: 10.1016/j.ab.2012.02.010.

[70] S. Varghese, F. Yang, V. Pacheco, K. Wrede, A. Medvedev, H. Ogata, M. Knipp, and H. Heise. Expression, Purification and Solid-state NMR Characterization of the Membrane Binding Heme Protein Nitrophorin 7 in two Electronic Spin States. *Biochemistry*, September 2013. doi: 10.1021/bi401020t.

[71] A. Oliveira, A. Allegri, A. Bidon-Chanal, M. Knipp, A.E. Roitberg, S. Abbruzzetti, C. Viappiani, and F.J. Luque. Kinetics and computational studies of ligand migration in Nitrophorin 7 and its Δ1-3 mutant. *Biochim. Biophys. Acta*, 1834(9):1711–21, September 2013. doi: 10.1016/j.bbapap.2013.04.009.

[72] A.R. Leach. *Molecular Modelling — Principles and Applications*. Prentice-Hall, New York-London-Toronto-Sydney-Tokyo-Singapore, 2001.

[73] T. Schlick. *Molecular Modeling and Simulation. An interdisciplinary Guide*. Springer-Verlag, 2002.

[74] P.K. Weiner and P.A. Kollman. AMBER: Assisted model building with energy refinement. A general program for modeling molecules and their interactions. *J. Comput. Chem.*, 2(3):287–303, 1981. ISSN 1096-987X. doi: 10.1002/jcc.540020311.

[75] B.R. Brooks, R.E. Bruccoleri, B.D. Olafson, D.J. States, S. Swaminathan, and M. Karplus. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.*, 4(2):187–217, 1983. ISSN 1096-987X. doi: 10.1002/jcc.540040211.

[76] J.C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R.D. Skeel, L. Kalé, and K. Schulten. Scalable Molecular Dynamics with NAMD. *J. Comput. Chem.*, 26(16):1781–802, December 2005. doi: 10.1002/jcc.20289.

[77] A. Bidon-Chanal. *Comprehensive study on Ligand diffusion in hemoproteins and receptor-ligand interactions.* PhD thesis, 2008.

[78] A.D. MacKerell, D. Bashford, M. Bellott, R.L. Dunbrack, J.D. Evanseck, M.J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F.T.K. Lau, C. Mattos, S. Michnick, T. Ngo, D.T. Nguyen, B. Prodhom, W.E. Reiher, B. Roux, M. Schlenkrich, J.C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin, and M. Karplus. All-atom empirical potential for molecular modeling and

dynamics studies of proteins. *J. Phys. Chem. B*, 102(18):3586–3616, 1998. doi: 10.1021/jp973084f.

[79] B.J. Alder and T.E. Wainwright. Phase Transition for a Hard Sphere System. *J. Chem. Phys.*, 27(5):1208–1209, 1957. doi: 10.1063/1.1743957.

[80] D.A. Case, T.A. Darden, T.E. Cheatham, C. Simmerling, J. Wang, R.E. Duke, R. Luo, R.C. Walker, W. Zhang, K.M. Merz, B. Roberts, B. Wang, S. Hayik, A.E. Roitberg, G. Seabra, I. Kolossváry, K.F. Wong, F. Paesani, J. Vanicek, J. Liu, X. Wu, S.R. Brozell, T. Steinbrecher, H. Gohlke, Q. Cai, X. Ye, J. Wang, M.-J. Hsieh, G. Cui, D.R. Roe, D.H. Mathews, M.G. Seetin, C. Sagui, V. Babin, T. Luchko, S. Gusarov, A. Kovalenko, and P.A. Kollman. *Amber 11*. University of California, San Francisco, 2010.

[81] J.P. Ryckaert, G. Ciccotti, and H.J.C. Berendsen. Numerical integration of the cartesian equations of motion of a system with constraints: Molecular Dynamics of $n-$alkanes. *J. Comp. Phys.*, 23:327–341, 1977. doi: 10.1016/0021-9991(77)90098-5.

[82] S. Miyamoto and P.A. Kollman. SETTLE: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. *J. Comput. Chem.*, 13(8):952–962, 1992. doi: 10.1002/jcc.540130805.

[83] H.J.C. Berendsen, J.P.M. Postma, W.F. van Gunsteren, A. Di Nola, and J.R. Haak. Molecular Dynamics with Coupling to an External Bath. *J. Chem. Phys.*, 81(8):3684–3690, 1984. doi: 10.1063/1.448118.

[84] W.G. Hoover. Canonical dynamics: Equilibrium phase-space distributions. *Phys. Rev. A*, 31(3):1695–1697, March 1985. ISSN 1050-2947. doi: 10.1103/PhysRevA.31.1695.

[85] S. Nosé. A unified formulation of the constant temperature molecular-dynamics methods. *J. Chem. Phys.*, 81(1):511–519, 1984. doi: 10.1063/1.447334.

[86] S.E. Feller, Y. Zhang, R.W. Pastor, and B.R. Brooks. Constant pressure molecular dynamics simulation: The Langevin piston method. *J. Chem. Phys.*, 103(11):4613–4621, 1995. doi: 10.1063/1.470648.

[87] P.P. Ewald. Die Berechnung optischer und elektrostatischer Gitterpotentiale. *Annalen der Physik*, 369(3):253–287, 1921. ISSN 1521-3889. doi: 10.1002/andp.19213690304.

[88] D.M. York, T.A. Darden, and L.G. Pedersen. The effect of long-range electrostatic interactions in simulations of macromolecular crystals: a comparison of the Ewald and truncated list methods. *J. Chem. Phys.*, 99:8345–8348, 1993. doi: 10.1063/1.465608.

[89] T.A. Darden, D.M. York, and L.G. Pedersen. Particle mesh Ewald — an $N \log N$ method for Ewald sums in large systems. *J. Chem. Phys.*, 98: 10089–10092, June 1993. ISSN 0021-9606. doi: 10.1063/1.464397.

[90] W. Humphrey, A. Dalke, and K. Schulten. VMD: visual molecular dynamics. *J. Mol. Graph.*, 14(1):33–8, 27–8, February 1996. ISSN 0263-7855. doi: 10.1016/0263-7855(96)00018-5.

[91] P. Schmidtke, A. Bidon-Chanal, F.J. Luque, and X. Barril. MDpocket: open-source cavity detection and characterization on Molecular Dynamics trajectories. *Bioinformatics*, 27(23):3276–85, December 2011. doi: 10.1093/bioinformatics.

[92] V. Le Guilloux, P. Schmidtke, and P. Tufféry. Fpocket: An open source platform for ligand pocket detection. *BMC Bioinformatics*, 10, 2009. doi: 10.1186/1471-2105-10-168.

[93] J. Cohen, A. Arkhipov, R. Braun, and K. Schulten. Imaging the migration pathways for O_2, CO, NO, and Xe inside Myoglobin. *Biophys. J.*, 91(5): 1844–57, September 2006. doi: 10.1529/biophysj.106.085746.

[94] J. Cohen and K. Schulten. O_2 migration pathways are not conserved across proteins of a similar fold. *Biophys. J.*, 93(10):3591–600, November 2007. doi: 10.1529/biophysj.107.108712.

[95] J. Cohen, K.W. Olsen, and K. Schulten. Finding gas migration pathways in proteins using implicit ligand sampling. *Methods Enzymol.*, 437:439–57, 2008. ISSN 0076-6879. doi: 10.1016/S0076-6879(07)37022-5.

[96] R.M. Daniel, R.V. Dunn, J.L. Finney, and J.C. Smith. The role of dynamics in enzyme activity. *Annu. Rev. Biophys. Biomol. Struct.*, 32(1):69–92, 2003. doi: 10.1146/annurev.biophys.32.110601.142445. PMID: 12471064.

[97] E.Z. Eisenmesser, D.A. Bosco, M. Akke, and D. Kern. Enzyme dynamics during catalysis. *Science*, 295(5559):1520–3, February 2002. ISSN 1095-9203. doi: 10.1126/science.1066176.

[98] K. Hinsen, A. Thomas, and M.J. Field. Analysis of domain motions in large proteins. *Proteins: Structure, Function, and Bioinformatics*, 34(3):369–382, 1999. ISSN 1097-0134. doi: 10.1002/(SICI)1097-0134(19990215)34:3.

[99] J. Luo and T.C. Bruice. Anticorrelated motions as a driving force in enzyme catalysis: the dehydrogenase reaction. *Proc. Natl. Acad. Sci. U.S.A.*, 101(36):13152–6, September 2004. ISSN 0027-8424. doi: 10.1073/pnas.0405502101.

[100] J. Ma and M. Karplus. The allosteric mechanism of the chaperonin GroEL: a dynamic analysis. *Proc. Natl. Acad. Sci. U.S.A.*, 95(15):8502–7, July 1998. ISSN 0027-8424. URL `http://www.ncbi.nlm.nih.gov/pubmed/9671707`.

[101] S. Sacquin-Mora and R. Lavery. Investigating the local flexibility of functional residues in hemoproteins. *Biophys. J.*, 90(8):2706–17, April 2006. ISSN 0006-3495. doi: 10.1529/biophysj.105.074997.

[102] L.-W. Yang and I. Bahar. Coupling between catalytic site and collective dynamics: a requirement for mechanochemical activity of enzymes. *Structure*, 13(6):893–904, June 2005. ISSN 0969-2126. doi: 10.1016/j.str.2005.03.015.

[103] A. Amadei, A.B. Linssen, and H.J.C. Berendsen. Essential dynamics of proteins. *Proteins*, 17(4):412–25, December 1993. ISSN 0887-3585. doi: 10.1002/prot.340170408.

[104] B. Hess. Similarities between principal components of protein dynamics and random diffusion. *Phys. Rev. E*, 62:8438–8448, Dec 2000. doi: 10.1103/PhysRevE.62.8438.

[105] B. Hess. Convergence of sampling in protein simulations. *Phys. Rev. E*, 65:031910, Mar 2002. doi: 10.1103/PhysRevE.65.031910.

[106] A. Pérez, J.R. Blas, M. Rueda, J.M. López-Bes, X. de la Cruz, and M. Orozco. Exploring the Essential Dynamics of B-DNA. *J. Chem. Theory Comput.*, 1(5):790–800, 2005. doi: 10.1021/ct050051s.

[107] N.A. Baker, D. Sept, S. Joseph, M.J. Holst, and J.A. McCammon. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. U.S.A.*, 98(18):10037–41, August 2001. ISSN 0027-8424. doi: 10.1073/pnas.181342398.

[108] V. Hornak, R. Abel, A. Okur, B. Strockbine, A.E. Roitberg, and C. Simmerling. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins*, 65(3):712–25, November 2006. doi: 10.1002/prot.21123.

[109] L. Boechi, P.A. Mañez, F.J. Luque, M.A. Martí, and D.A. Estrin. Unraveling the molecular basis for ligand binding in truncated Hemoglobins: the trHbO *Bacillus subtilis* case. *Proteins*, 78(4):962–70, March 2010. doi: 10.1002/prot.22620.

[110] J. Muñoz-Muriedas, X. Barril, J.M. López, M. Orozco, and F.J. Luque. A hydrophobic similarity analysis of solvation effects on nucleic acid bases. *J. Mol. Model.*, 13(2):357–65, February 2007. doi: 10.1007/s00894-006-0150-y.

[111] A. Bender and R.C. Glen. Molecular similarity: a key technique in molecular informatics. *Org. Biomol. Chem.*, 2(22):3204–18, November 2004. doi: 10.1039/B409813G.

[112] J.B. Klauda, R.M. Venable, J.A. Freites, J.W. O'Connor, D.J. Tobias, C. Mondragon-Ramirez, I. Vorobyov, A.D. MacKerell, and R.W. Pastor. Update of the CHARMM all-atom additive force field for lipids: validation on six lipid types. *J. Phys. Chem. B*, 114(23):7830–43, June 2010. ISSN 1520-5207. doi: 10.1021/jp101759q.

[113] S. Jo, T. Kim, and W. Im. Automated builder and database of protein/membrane complexes for molecular dynamics simulations. *PLoS ONE*, 2(9):e880, 2007. doi: 10.1371/journal.pone.0000880.

[114] S. Jo, J.B. Lim, J.B. Klauda, and W. Im. CHARMM-GUI Membrane Builder for mixed bilayers and its application to yeast membranes. *Biophys. J.*, 97(1):50–8, July 2009. doi: 10.1016/j.bpj.2009.04.013.

[115] S. Jo, T. Kim, V.G. Iyer, and W. Im. CHARMM-GUI: a web-based graphical user interface for CHARMM. *J. Comput. Chem.*, 29(11):1859–65, August 2008. doi: 10.1002/jcc.20945.

[116] S.A. Adcock and J.A. McCammon. Molecular Dynamics: survey of methods for simulating the activity of proteins. *Chem. Rev.*, 106(5):1589–615, May 2006. doi: 10.1021/cr040426m.

[117] W.L. Jorgensen, J. Chandrasekhar, J.D. Madura, R.W. Impey, and M.L. Klein. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.*, 79:926–935, 1983. doi: 10.1063/1.445869.

[118] Schrödinger, LLC. The PyMOL Molecular Graphics System, Version 1.3r1. August 2010. URL `http://pymol.org`.

[119] J.D. Thompson, T.J. Gibson, F. Plewniak, F. Jeanmougin, and D.G. Higgins. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.*, 25(24):4876–82, December 1997. ISSN 0305-1048. doi: 10.1093/nar/25.24.4876.

Appendix A

MD settings with NAMD/charmm

A.1 A typical setup

Typically, for NP7 in solution, a standard NAMD/charmm [76, 78] steps and setup for the different pre-production and production dynamics runs are summarized in Tables A.1 and A.2.

A.1.1 Preparation

Given an initial structure for NP7, the preparatory steps before submitting MD calculations include:

1. parsing the pdb file checking residue names, protonation states for histidines, heme orientation;

2. building the psf structure file; to this end, the *AutoPSFbuilder* extension of VMD [90], loading the standard charmm22 topology files, including also the specific heme and CO topology files [78], specifying the patches (ASPP if Asp32 is to be protonated, DISU for disulfide bridges, PHEM for the His60-Heme coordination and PLIG for the CO-Heme coordination);

3. building the water box around the protein, a 69 Å cubic box in order to get approximately a system with 30k atoms, as in previous Amber simulations; this was done using the VMD *Solvate* extension; the water model is TIP3P [117];

4. introducing, as substituents to water molecules, counterions in order to have an overall neutral box; this was done using the VMD *Autoionize* extension.

step	type	constraints & conditions	length
min1	energy minim.[d]	non-water H[b]	2000 steps
min2	energy minim.[d]	H_2O[b]	5000 steps
min3	energy minim.[d]	all atoms[b]	10000 steps
eq1	equilibration	NVT, T=100 K$\xrightarrow{50\,K}$300 K	50000 steps[a]
eq2	equilibration	NPT, T=300 K, p=1.0 atm	100000 steps[a]
mdx	production dynamics	NPT, T=300 K, p=1.0 atm	n steps[a,c]

Table A.1: NAMD typical pre-production and production segments division for NP7 in solution. [a] The timestep for equilibration and production dynamics amounts to a simulated time of 2 fs, see Table A.2. [b] Atoms hereby indicated are those unconstrained. [c] Production dynamics runs span typically in the range of hundreds of ns, and are split in shorter segments according to the computing facility's queuing policies. [d] Conjugate gradient energy minimization.

The result is a pdb-psf pair of files that, together with the files containing the standard charmm22 parametres (with cmap corrections), are all is needed to start the simulations.

A.1.2 Energy minimization

Common settings to all NAMD runs, independently on the particular step, include the Particle mesh Ewald method for the evaluation of long-range electrostatic interactions [88, 89], the presence of a cutoff for the evaluation of all non-bonded (i.e. Van der Waals and coulombic) interactions and the use of the SHAKE algorithm [81] in order to keep all bonds involving hydrogen atoms at their equilibrium length.

The first three steps of the MD sequence are meant to make the system more statically realistic by minimizing the system energy upon small atomic displacements.

Since the hydrogen atoms, water molecules and counterions were automatically placed with the only constraints of correct quantities and absence of steric clashes, the energy minimization focuses initially on these sets of atoms: non-water hydrogen atoms and then water molecules and ions. Only in a third time, all the atoms are unconstrained and are considered by the NAMD conjugate gradient algorithm.

The dimensions of the periodic-boundary box are specified only in the first energy minimization step, all subsequent steps gathering information from the output of the previous step, as since the second equilibration step the volume will be left free to change.

variable	value
cell shape and size	cubic, $l = 69\,\text{Å}$[a]
timestep	2 fs
snapshots	1 ps
boundary conditions	periodic
electrostatics	Particle Mesh Ewald, grid spacing 1 Å
constraints	on[b], harmonic
non-bonded cutoff	14 Å
non-bonded switching	on[c], distance 9 Å
water model	TIP3P
ShakeH	all H-X bonds rigid, tolerance 1×10^{-5} Å
Langevin dynamics	on[d] heavy atoms, T=300 K, damping 5 ps^{-1}
Nosé-Hoover Langevin piston	on[e], P=1 atm, period 100 fs, decay 50 fs

Table A.2: NAMD typical pre-production and production settings for NP7 in solution. [a] Only specified in the min1 step (see Table A.1), subsequently taken from last output of previous step. [b] Position constraints are on only during min1 and min2 (see Table A.1). [c] Only active in equilibration and production dynamics. [d] Only active during production dynamics. [e] Only active during eq2 and production dynamics.

A.1.3 Equilibration

The next two steps are meant to add dynamical realism to the system. Common variables to equilibration and production dynamics include the size of the timestep used for the integration of Newton's equations, 2 fs, a value that is compatible with the use of the SHAKE algorithm.

The first equilibration phase is thermalization. Velocities are pseudorandomly assigned to all atoms in order to simulate a low initial temperature. These velocities are then repeatedly modified after a certain simulation time in order to increase the simulation temperature and allow a gradual relaxation and adaptation of the protein structure to the new thermal motions. At the end of this phase, the final 300 K simulation temperature is reached and some extra time is allowed for further structural adjustments.

The second equilibration phase is pressurization. A Nosé-Hoover Langevin, thermally-coupled piston is switched on with a target pressure of 1.0 atm at 300 K, while the box dimensions are kept free to change as before.

Since the equilibration runs, an additional switching function is added to smooth the non-bonded cutoff in order to avoid discontinuities in the potential.

variable	value
α-sphere min radius	2.8 Å
α-sphere max radius	5.5 Å
number of snapshots per traj. segment	5000
number of traj. segments	≥ 2

Table A.3: MDpocket typical settings for NP7 in solution.

variable	value
resolution	0.2 Å
subsampling	3
orientations	20
cutoff energy	85
temperature	300 K
nonbonded cutoff	12.0 Å
probe CO: VdW radius	2.1 Å / 1.7 Å
probe CO: VdW epsilon	-0.11 / -0.12
probe CO: charge	+0.021e / -0.021e
number of snapshots per traj. segment	5000
number of traj. segments	≥ 2

Table A.4: ILS VMD plugin typical settings for NP7 in solution.

A.1.4 Production dynamics

The production dynamics runs can then start, with an additional Langevin temperature control.

A.2 Auxiliary techniques

In Tables A.3, A.4 and A.5 the common settings for MDpocket (see 2.4.2), ILS (see 2.4.3) and APBS (see 2.4.5), respectively, are summarized.

variable	value
grid points per processor	129/161
coarse grid lengths	approx. $90 \times 80 \times 70$Å
coarse grid lengths	approx. $75 \times 65 \times 60$Å
PB equation	linearized
boundary conditions	single Debye-Hückel
dielectric constant (protein)	20.00
dielectric constant (solvent)	78.50
dielectric ion-accessibility coefficients	smol[a]
grid points density in discont. surfaces	40.00 pt·Å^{-2}
charge mapping	cubic B-spline
solvent radius	1.40 Å
spline window	0.30 Å
temperature	300 K

Table A.5: APBS typical settings as used in Chapter 4. [a] See also on http://www.poissonboltzmann.org the APBS user guide for a complete description.

Appendix B

Software

Visualization and graphics: VMD[90], grace, qtiplot, PyMOL[118];

Modeling: VMD (see above), CHARMM-GUI (http://www.charmm-gui.org) [115], CHARMM-GUI Membrane Builder[113, 114];

MD: NAMD[76], charmm22[78];

Poisson-Boltzmann equations: APBS[107] and pdb2pqr;

sequence alignment: Clustal X[119], Treeview X;

for editing the present document: KBibTEX, pdfLaTeX.