In silico investigation of the molecular effects caused by R123H variant in secretory phospholipase A2-IIA associated with ARDS

Benedetta Righino[a], Angelo Minucci[a,b], Davide Pirolli[e], Ettore Capoluongo[a,b], Giorgio Conti[c], Daniele De Luca[c,d] and Maria Cristina De Rosa[e*]

[a] Institute of Biochemistry and Clinical Biochemistry, Catholic University of the Sacred Heart, L.go F. Vito 1, 00168 Rome, Italy

[b] Department of Laboratory Medicine, "Policlinico Gemelli" Foundation, L.go A. Gemelli 1, 00168 Rome, Italy

[c] PICU, Department of Anesthesiology and Critical Care, Catholic University of the Sacred Heart, L.go F. Vito 1, 00168 Rome, Italy

[d] APHP, South Paris University Hospitals, Medical Center "A. Beclere", Paris, France

[e] Institute of Chemistry of Molecular Recognition - CNR, 00168 Rome, Italy

*Corresponding author:
Maria Cristina De Rosa
Institute of Chemistry of Molecular Recognition - CNR, 00168 Rome, Italy c/o Catholic University of the Sacred Heart, L.go F. Vito 1, 00168 Rome, Italy
Email: mariacristina.derosa@icrm.cnr.it
phone: +39 06 30155135

Abstract

Phospholipase A2-IIA catalyzes the hydrolysis of the sn-2 ester of glycerophospholipids. A rare c.428G>A (p.Arg143His) variant in PLA2G2A gene was found in two infants affected by acute respiratory distress syndrome (ARDS) by whole coding region and exon/intron boundaries sequencing. To obtain insights into the possible molecular effects of the rare R123H mutation in secretory PLA2-IIA (sPLA2-IIA), molecular modelling, molecular dynamics (MD) using principal component analysis (PCA) and continuum electrostatic calculations were conducted on the crystal structure of the wild type protein and on a generated model structure of the R123H mutant. Analysis of MD trajectories indicate that the overall stability of the protein is not affected by this mutation but nevertheless the catalytically crucial H-bond between Tyr51 and Asp91 as well as main electrostatic interactions in the region close to the mutation site are altered. PCA results indicate that the R123H replacement alter the internal molecular motions of the enzyme and that collective motions are increased. Electrostatic surface potential studies suggest that after mutation the interfacial binding to anionic phospholipid membranes and anionic proteins may be changed. The strengthening of electrostatic interactions may be propagated into the active site region thus potentially affecting the substrate recognition and enzymatic activity. Our findings provide the basis for further investigation and advances our understanding of the effects of mutations on sPLA2 structure and function.

Keywords: molecular dynamics simulation; sPLA2; PLA2G2A gene; molecular modelling; electrostatic potential

1. Introduction

Phospholipases A2 (phosphatide 2-acylhydrolase, EC 3.1.1.4) are a widely distributed group of enzymes that catalyse the hydrolysis of ester bonds at the *sn*-2 position of phospholipids [1]. Secretory phospholipases A2 (sPLA2) represent the extracellular forms secreted by several cell types; over 10 distinct sPLA2 isotypes have been described in mammalians and the -IIA (sPLA2-IIA) is the main pulmonary isotype, secreted into the alveoli by alveolar macrophages [2].

sPLA2s are relevant in lung physiopathology, since they may contribute to the production of inflammatory mediators and are responsible for surfactant phospholipids catabolism. These two processes are typical of acute respiratory distress syndrome (ARDS), which is a life-threatening condition characterized by high lung tissue inflammation and surfactant dysfunction [1,3]. sPLA2-IIA is increased in broncho-alveolar lavages (BAL) from ARDS animal models [4,5]; moreover BAL levels of sPLA2-IIA are increased in ARDS adult patients and correlate with clinical severity[6,7]. Given the role of sPLA2 in ARDS in adult patients, we set up an international network to investigate the role of sPLA2 in various pediatric critical respiratory conditions, designing a multi-step translational project whose plan has been described elsewhere [8]. We demonstrated that sPLA2-IIA is a main enzyme isotype in infants with post-neonatal ARDS: moreover, in these patients, sPLA2 activity correlates with clinical severity and also with mortality, lenght of pediatric intensive care (PICU) stay and respiratory outcomes [9,10]. Pediatric ARDS is defined as an acute-onset multifactorial condition needing peculiar triggering factors and an underlying given degree of predisposition [11]. In fact, polymorphisms of the NFκB gene, which is the sPLA2-IIA transcription factor, are associated with the risk to develop ARDS and its outcomes [12,13]. It is known that a sPLA2 gene polymorphism is associated with complications of chronic obstructive pulmonary disease, however data about sPLA2 genetics and ARDS are lacking [14]. Within our multi-step project we also meant to study sPLA2 genetics, thus we sequenced sPLA2-IIA gene (PLA2G2A) in infants with ARDS [8].

sPLA2-IIA protein has been described in detail [1]. It is characterized by a low molecular weight (14 kDa), seven conserved disulfide bonds and His/Asp dyad for the catalytic mechanism and Ca^{2+} bound in the active site. The active site possesses a hydrophobic channel, which is formed by Leu2, Val3, Phe5, His6, and Ile9 on the N-terminal α1 helix (2-13) and Alal7 and Ala18 on the short α2 helix (16-24). The calcium ion, which is essential for the sPLA2 -IIA activity is bound to the main chain carbonyl oxygen atoms of residues His28, Gly30, and Gly32 and to the side chain oxygen atoms of Asp48 at the active site which includes the catalytic His47 and Asp91. The catalytic mechanism for hydrolysis of ester bonds at the sn-2 position of phospholipids is shown in Fig. 1.

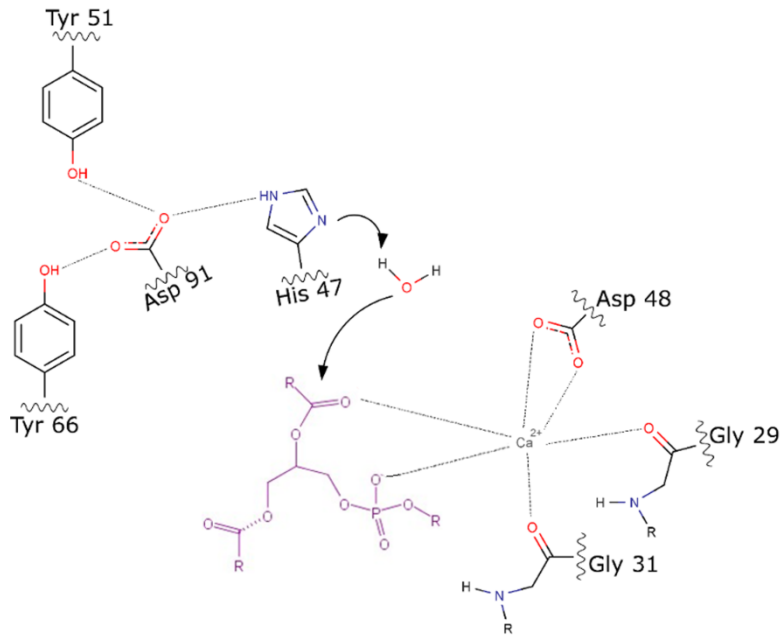

Figure 1. Schematic representation of the catalytic mechanism of sPLA. The catalytic site consists essentially of His47, Asp91, a calcium ion (cofactor), and a water molecule that acts as the nucleophile. The hydrogen atom of the water molecule is transferred to His47 and the calcium ion stabilizes the oxyanion derived from the substrate's carbonyl oxygen. The histidine imidazole ring becomes positively charged and the protonated His47 is stabilized by Asp91, which hydrogen-bonds to Tyr51 and Tyr66. Residue numbering is based on the PDB file 1DB5 [15]. Substrate is shown in magenta.

The crystal structure and surface electrostatic potential calculation of sPLA2-IIA have shown that this highly cationic protein contains a large number of cationic patches on its molecular surface [16-18]. Some cationic patches are essential for interfacial binding, whereas Arg123 belongs to a cluster of basic residues (cluster A, including Lys52, Arg53, Lys56 and Arg57) which is involved in heparinoid binding [19,20].

Here, we report the discovery of a rare variant in the PLA2G2A gene of two pediatric patients, identical twins, affected by ARDS. Predicting how this mutation alters the structure and function of sPLA2-IIA may be helpful for the identification of the role of sPLA2s in the mechanism underlying the disease. Molecular dynamics simulation has been widely used to study the effect of mutations on the structure and function of proteins [21-24]. Here we have used molecular modelling, molecular dynamics and electrostatic potential calculation to understand the conformational changes and dynamics of sPLA2 under the effect of R123H mutation. Our investigation provides detailed atomistic insights into the structural features of this rare variant.

2. Materials and methods

2.1. Population

Twenty-four infants diagnosed with ARDS according to the American-European Consensus Criteria were recruited [25]. These were patients previously enrolled in our earlier investigation step about biological and clinical effect of sPLA2 during pediatric ARDS and basic population details have been described elsewhere [10]. These patients were subjected to broncho-alveolar lavage and sPLA2 activity was measured in their epithelial lining fluids (ELF). Details about lavage procedure, sPLA2 activity assay and ELF estimation have been published elsewhere [10]. We also enrolled as controls 25 healthy age matched (within 2 months) infants admitted to our pediatric intensive care unit (PICU) in the month immediately before or after the ARDS infants. These controls were all admitted for postsurgical care, they had no lung disease in the previous three months and they have

never been diagnosed with ARDS before. During hospitalization they had a normal lung imaging and clinical examination and never needed any supplemental oxygen.

Finally, 50 non-smoker adults, never diagnosed with ARDS and with no lung disease in the previous three months, have been recruited as further control. These were outpatients seen for glucose-6-phosphate dehydrogenase deficiency genetic test. In all these control subjects 1 mL whole blood was drawn into EDTA tubes at the moment of blood drawings for clinical routine tests and kept at 4°C. No venepuncture was performed solely for study purposes and the participation to the study did not change patients' care; institutional review board approved the study and informed consent was asked to patients or parents/guardians.

2.2. Molecular testing of the PLA2G2A gene

Genomic DNA was isolated from peripheral blood by the MagCore® Genomic DNA Whole Blood Kit, according to the manufacturer's procedures and using the MagCore® Automated Nucleic Acid Extractor (Diatech Lab Line srl. Jesi, Italy). The published genomic sequence of the PLA2G2A gene and the NCBI Reference Sequence NG_012928.1 were used to design primers for PCR amplification of four genomic coding fragments also covering the exon-intron junctions: fragments 1, 2, 3 (primers: 1AF-1AR: 5'-AAGTTGAGACCACCCAGCAG-3', 5'-TTTTCCCCCTGAGAGAGGAT-3'; 2AF-2AR: 5'-CCATTTGGGAGGAAGGAGA-3', 5'-AGGAACCGGCACTGTCTTT-3'; 3AF-3AR: 5'-CTGGAGCTGTGGGACAAGA-3', 5'-CACAGTCCCCAGCACTGTCTA-3') encompassing exons 2, 3, 4 and fragment 4 (primers: 4AF-4AR: 5'-CCCACAAGAAGCCACTGAAT-3', 5'-AATTCAGCACTGGGTGGAAG-3') encompassing exon 5, were generated [26]. PCR products were sequenced using the Big Dye Terminator v3.1 Cycle Sequencing kit (Applied Biosystems, Foster City, CA, USA, http://www.appliedbiosystems.com/absite/us/en/home.html) in an automated sequencer ABI Prism 3500 Genetic Analyzer (Applied Biosystems). The same primers were used for PCR amplification

and direct sequencing. Using the SeqScape® Software v2.5, sequences were aligned to the reference sequence NG_012928.1. Allelic frequencies were compared with Fisher exact test.

2.3. Bioinformatics Analyses

The potential consequences of the mutation were first predicted by means of two bioinformatics methods: PolyPhen-2 [27] and SIFT [28] using default settings for mutation assessment. Obtained scores, 0.8 (possibly damaging) and 0.06 (tolerated) for PolyPhen-2 and SIFT respectively, suggest further investigations.

2.4. Mutation modelling

The structure of the SPLA2-IIA R123H mutant was generated by using the Build Mutant tool of the Protein Modelling module of Discovery Studio (Dassault Systèmes BIOVIA, Discovery Studio Modeling Environment, Release 2017, San Diego: Dassault Systèmes, 2016). We used the structure of human sPLA2-IIA in complex with indole 6 (available in the Protein Data Bank, PDB: 1DB5) as the starting point for molecular modelling studies [15]. This structure was chosen, among those available in the PDB, on the basis of the best stereochemical quality as calculated by PROCHECK [29] (93.6%, 6.4%, 0% and 0% the percentage of the residues belonging to the most favoured, additionally allowed, generously allowed and disallowed region of the plot, respectively). The Build Mutant protocol automatically mutates specific residues and optimizes the confirmation of neighbouring residues that lie within a specified cutoff radius of 4.5 Å [30]. The generated 50 model structures were then evaluated using the DOPE scoring function and the PDF total energy of the program. The structure with the lowest PDF (probability density function) was selected and checked using the programs PROCHECK, VERIFY3D and ProSA-Web (https://prosa.services.came.sbg.ac.at/prosa.php) [29,31,32].

2.5. Molecular dynamics simulation protocol

We used the structure of human s-PLA2-IIA (PDB: 1DB5) and the generated R123H mutant structure as the starting point for the MD simulations in water. Hydrogens were added to the crystal and modelled structures using the Maestro program (Schrodinger Inc.). Two independent simulations were set up for the wild and mutated structures. MD simulations were carried out using the program Desmond as implemented in Maestro (Schrödinger Release 2017-1: Desmond Molecular Dynamics System, D. E. Shaw Research, New York, NY, 2017. Maestro-Desmond Interoperability Tools, Schrödinger, New York, NY, 2017). Each simulated system was solvated in an orthorhombic box (53 Å x 46 Å x 67 Å and 53 Å x 44 Å x 67 Å) for wild type and R123H mutant, respectively) with explicit SPC water molecules (4311 and 4131, respectively). A 0.15 M NaCl salt concentration was added and 18 (wild type) and 17 (R123H mutant) chloride ions were added to the systems to maintain electric neutrality. Atomistic interactions were calculated with the OPLS 2005 force field [33]. The particle-mesh Ewald method was used to calculate the long-range electrostatic interactions. A cut-off radius of 9.0 Å was applied for short-range van der Waals and Coulomb interactions. The systems were heated up and equilibrated with the standard Desmond Maestro protocol and then simulated under an isothermal–isobaric ensemble (NPT) with the temperature of 300 K and the pressure of 1 bar. A Nose-Hoover thermostat and Martyna–Tobias–Klein method were implemented to maintain the temperature and the pressure of the systems, respectively [34,35]. A time step of 2.0 fs was used for the overall simulations. The systems were minimized and equilibrated with the default protocols of Desmond. 200 ns non-constrained MD simulation was performed for each system, and the coordinates were saved for every 200 ps. Different random number seeds were used for independent trajectories. Bio3D package for R was used to analyse motion of the Cα atoms by principal component analysis and to detect time correlated motions of all C-alpha atom pairs by dynamic cross-correlation maps [36].

2.6. Electrostatic Potential Calculations

The calculation for sPLA2-IIA and its R123H variant was based on the model structures generated from MD simulations. Electrostatic potentials and electrostatic interaction energies were calculated with the DelPhi software package, as implemented in Discovery Studio 4.5, which solves the nonlinear Poisson Boltzmann equation by the finite difference method [37]. The calculation employed a grid of 49x42x36 with a grid resolution of 0.7 Å. Dielectric constant of 80 and 2 were used for solvent and solute, respectively. An ionic exclusion radius of 2.0 Å, a solvent radius of 1.4 Å and a solvent ionic strength of 0.145 M were applied. The electrostatic potential distribution was mapped onto the solvent accessible surface calculated using Discovery Studio with the radius of the probe set at 1.4 Å. pKa values of the individual amino acid residues were calculated using the approach of Spassov and Yan as implemented in Discovery Studio 4.5 [38].

3. Results and discussion

3.1. Clinical and genetic data

Direct sequencing revealed no mutations in 22 patients, albeit a missense c.428G>A (rs34568801) variant in exon 5 of the PLA2G2A gene was found in 2 identical twins affected by ARDS. The variant causes the amino-acid change p.Arg123His in the mature form of the protein. To date, this variant is only reported in the dbSNP database (https://www.ncbi.nlm.nih.gov/projects/SNP/) and no clinical significance is associated. Its allelic frequency is significantly different between cases and control infants ($p=0.05$; Fisher test) and also between cases and healthy adult controls ($p=0.01$; Fisher test). Minor Allele Frequency (MAF) is also significantly different, as this variant is considered a rare one since it has a global MAF of 0.0020, while it is 0.023 in our population ($p=0.04$; Fisher test). Table 1 shows basic clinical characteristics of the two patients and of control patients admitted to pediatric intensive care unit for reasons other than ARDS. The two cases were affected by a moderately severe ARDS as seen by oxygenation and compliance measurements, while control patients did not have any derangement in oxygenation, compliance and different

PRISM III-24 score and clinical outcomes. The other control group consisted of outpatients tested for glucose-6-phosphate dehydrogenase deficiency and otherwise healthy, thus physiological and vital parameters were not measurable for them and were not reported in Table 1, as they did not need any critical care monitoring. The sPLA2 activity in ELF at the ARDS diagnosis resulted also moderately high, as it was 210 IU/mL and 312 IU/mL, for case 1 and 2, respectively. The two subjects carrying the mutated PLA2G2A allele were identical twins. Gene sequencing of the patients' parents showed that the variant was of paternal origin. No episodes of ARDS were reported in the family history. Patients have been followed up for 6 years: they never presented new ARDS episodes and their growth, development and clinical history has always been normal. Our findings align with those indicating that PLA2G2A SNP is potentially associated with neonatal sepsis [39,40].

Table 1

Clinical data of the two patients and of the control group of infants admitted to pediatric intensive care unit without ARDS.

	Case 1	**Case 2**	**Controls**
Age (months)	3	3	4.6 (0.1)
Male Sex	Yes	Yes	13 (52%)
PRISM III-24 score	22	25	5.4 (2.3)
PaO_2/FiO_2 ratio at the ARDS	95	79	490 (157)
OI at the ARDS diagnosis	18	29	1 (0.7)
Crs (mL/cmH_2O/kg) at the	0.28	0.4	0.89 (0.11)
PICU stay (days)	6	6	11.8 (9.6)
Duration of mechanical	5	5	9.2 (9.4)
Duration of O_2 therapy (days)	6	6	9.4 (9.4)

Abbreviations: PRISM III-24: Pediatric risk index for mortality, calculated as the worst in the first 24h from the PICU admission; FiO_2: inspired oxygen fraction; OI: oxygenation index, calculated as inspired oxygen fraction (FiO_2) x mean airway pressure / PaO_2 (dimensionless number); PaO_2: arterial partial tension of oxygen; Crs: static compliance of the respiratory system; PICU: pediatric intensive care unit; (more details in the text).Control group data are expressed as mean (standard deviation) or number (%), as appropriate.

3.2. Modelling of R123H sPLA2-IIA mutant

To investigate the effect of R123H mutation, homology modelling was used to construct the model of the mutant based on the crystal structure of sPLA2-IIA (PDB code: 1DB5). The overall stereochemical quality of the structures was evaluated using PROCHECK which bases its evaluation on a database of well refined high-resolution protein structures and presents the results as Ramachandran plots of residue phi-psi torsion angles. From the analysis of the crystal structures the program divides the Ramachandran space into four regions (most favourable, additionally allowed, generously allowed and disallowed) and a good quality model would be expected to have over 90% of phi-psi angles in the most favourable regions. The starting structure displayed 93.6% of the residues in the most favorable region, 6.4% in the additional allowed region and 0.0% in generously allowed and disallowed regions The fifty different generated models were carefully analyzed for energy value and probability density function (PDF) violations. The model with the lowest PDF value was selected as the representative model for R123H sPLA2-IIA structure. The reliability of the mutant structure was also checked with VERIFY-3D, which evaluates the compatibility of a given residue in a certain three-dimensional environment. A score below zero for a given residue means that the conformation adopted by that residue in the model is not compatible with its surrounding environment. The VERIFY-3D scores of our models are always positive. In order to verify whether the interaction energy of each residue with the remainder of the protein is negative (typically positive values correspond to tricky parts of a model) the protein structures were submitted to the ProSA web server available at https://prosa.services.came.sbg.ac.at/prosa.php. The low ProSA z-score value obtained (-5.6), within the range of scores typically found for native proteins of similar size, confirms the good quality of our model and the ProSA analysis of the mutant model shows that almost all the residues have negative values of interaction energy (Supplementary Information Fig. S1).

3.3. Molecular dynamics simulations

3.3.1. Protein stability

In order to identify the structural-dynamical properties that might suggest the alteration of enzymatic activity, comparative MD simulations of wild type (WT) and R123H mutant sPLA2-IIA have been carried out. Two independent simulations of 200 ns each have been performed which gave similar results. Figure 2 shows the root mean square deviation (RMSD), root mean square fluctuation (RMSF), solvent accessible surface area (SASA) and radius of gyration (Rg) of one of the simulation pairs, discussed here. Plots of the replicates are shown in the Supplementary Information (FigS2). The Cα-RMSD with respect to the model is shown in Fig. 2A for the WT simulation along with that for R123H mutant. RMSD values of the Cα atoms with respect to the initial structure indicates that within 1.5 ns a stable behaviour is reached, showing that the WT and R123H mutant proteins are equilibrated. For both simulations, the protein remained stable throughout, indicating that the conformations of WT and mutant protein are in good equilibrium. In WT sPLA2-IIA the Cα-RMSD exhibits a bit lower conformational variability compared to R123H sPLA2-IIA (average Cα-RMSD are 1.53 ± 0.19 Å and 1.71 ± 0.27 Å, respectively). The differences in the mobility of residues can be further investigated from the values of RMSF relative to the initial structure and are shown in Fig. 2B. The RMSF analysis for WT and mutant sPLA2-IIA indicate that the two proteins exhibit similar dynamical behaviors. However regions 48-63, 72-73 and 102-109 are more fluctuating in the mutant protein when compared to the wild type one. RMSF values fluctuate within a range of 0.40-2.7 Å (WT) and 0.50-2.90 Å (mutant) over the entire simulation period (Fig. 2B). We also analyzed the compactness of the enzyme by means of radius of gyration (Rg) and solvent accessible surface area (SASA). The average values of radius of gyration for WT and R123H sPLA2-IIA are 14.60 ± 0.11 Å and 15.02 ± 0.17 Å, respectively (Fig. 2C). These values indicate that the two proteins have similar globularity. Analysis of trajectories also indicates that the WT and mutant sPLA2-IIA have quite similar accessible area throughout the MD simulation, with SASA average values of 7972 ± 139 $Å^2$ and 8213 ± 211 $Å^2$, respectively (Fig. 2D). On the whole, the above analysis indicate that the system remains stable throughout the

simulation time upon R123H mutation; the similarities between the radius of gyration and the residue surface area, suggest WT and R123H sPLA2-IIA show similar global properties.

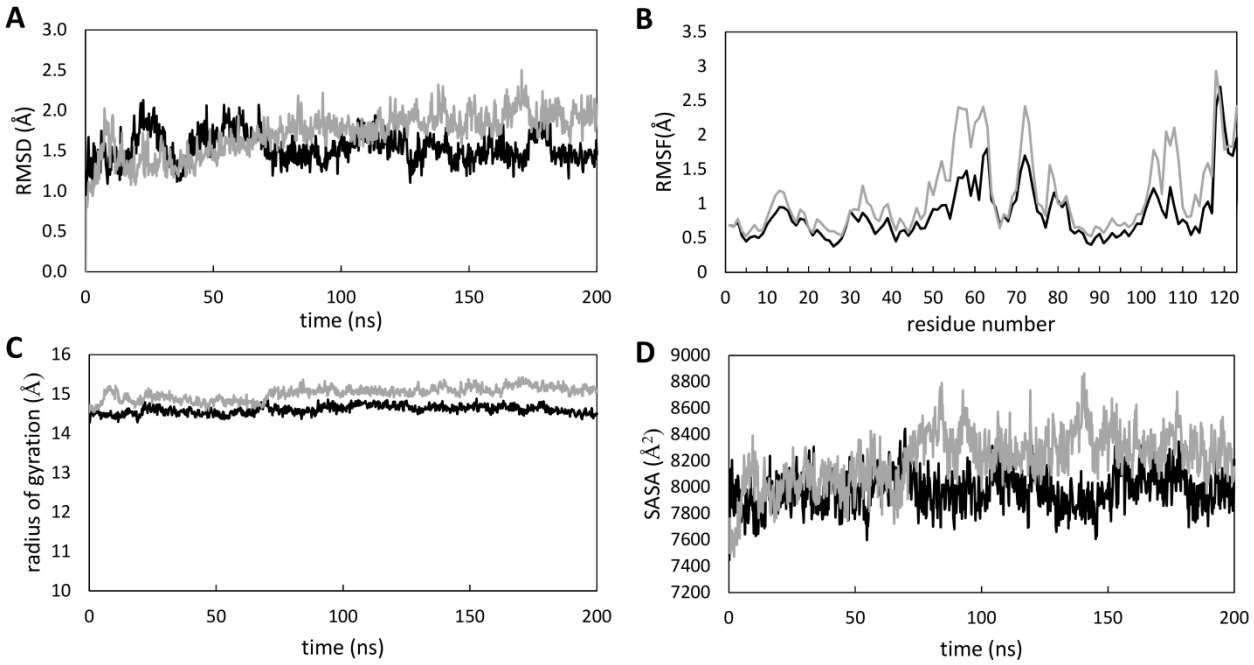

Figure 2. Evolution of structural properties over time. (A) Cα-RMSD as a function of time. (B) RMSF as a function of amino acids. (C) Radius of gyration as a function of time. (D) Solvent accessible surface area (SASA) as a function of time. Black-WT, Grey-R123H.

Besides the presence of seven disulfide bridges, the conformation of sPLA2-IIA is further constrained by the existence of α-helices leading to a 40% helical conformation of the secondary structure. For the above reasons the resulting MD trajectories were also analyzed by monitoring how secondary structure elements and significant distances varied during the duration of the simulations. First, we plot the time evolution of the secondary structure of the WT and mutant sPLA2-IIA from the MD (Fig. 3). In the WT run all secondary structure elements of sPLA2-IIA can be considered as stable (Fig. 3A, B). Notably, the N-terminal helical conformation (residues 1-7 and 15-20) which is lost during the equilibration phase, refolds at 23 ns and 5 ns, respectively (Fig. 3A). Focusing on the mutant structure, the helix spanning residues 39-56 partially unfolds and R123H mutation leads to a 40% decrease in helical content (Fig. 3C, D). Here, residues Cys49, Cys50 and

Tyr51 persistently adopt loop (white) conformation (Fig. 3C, D). Plots of the replicates are shown in the Supplementary Information, FigS3.

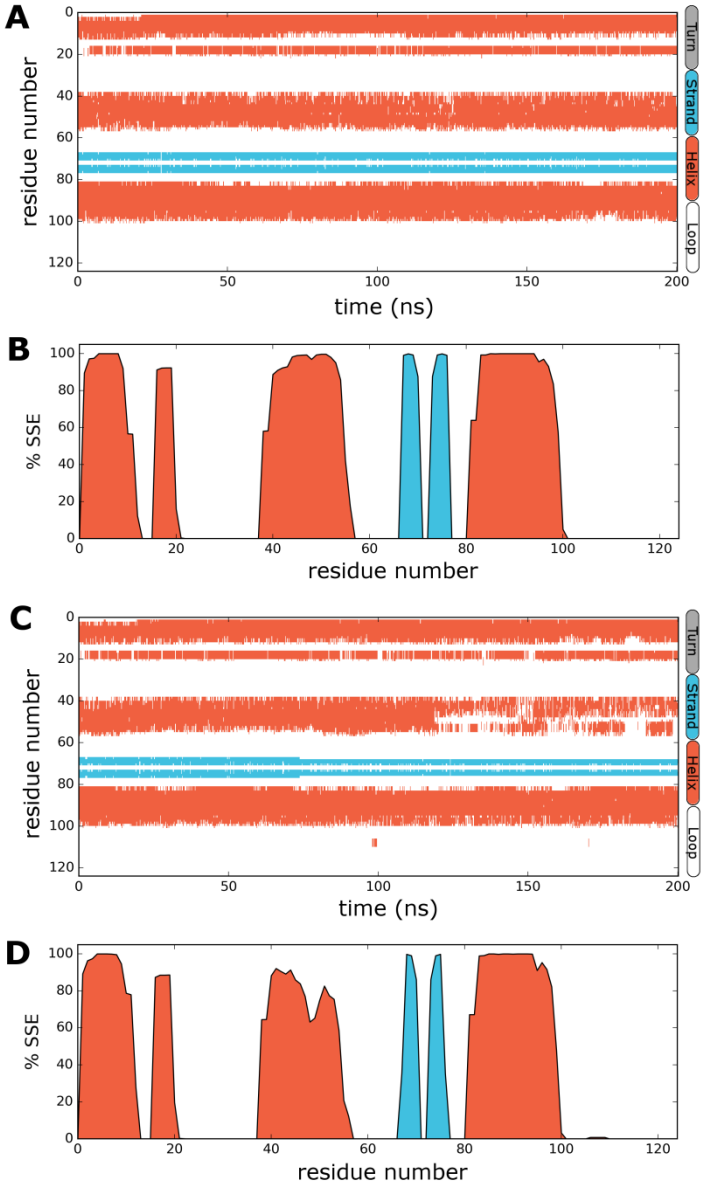

Figure 3. Graphical representation of secondary structure analysis. The secondary structure content of the protein as a function of time and the percentage of time each residue contributes to each of the three secondary structure elements: strand, helix, and loop for WT (A, B) and R123H mutant (C, D).

3.3.2. Principal component analysis and dynamic cross-correlation maps

We have analyzed the MD trajectory by principal component analysis (PCA), using the Cα atoms, to extract the significant collective motions showed by wild-type and mutant sPLA. The

overall flexibility was analyzed by the trace of the diagonalized covariance matrix. The trace value for WT and R123H structures was found to be 82.25 Å^2 and 136.87 Å^2 respectively, indicating that the mutant showed significant increased flexibility in the collective motion as compared to the WT. Fig. 4A shows the eigenvalues plotted against the corresponding eigenvector obtained from the diagonalization of the covariance matrix of the atomic fluctuations. The first few eigenvalues are relative to concerted motions and quickly reduce in amplitude to reach more localized fluctuations. The results indicate that the first ten principal components (PC) account for 65% (WT) and 73% (mutant) of the motions observed in the last 40 ns of the trajectories. It was also observed (Fig. 4A) that the overall motion between the two systems is varied. The magnitude of PC1 is increased by the mutation. In order to understand how R123H mutation affects the motions described by PC1, the displacements of PC1 for both structures were calculated. Fig. 4B suggests that R123H mutation can influence the displacements of the component in the regions 46-52, 55-63 and 123-124. This indicates that R123H can have influence on the motions of the basic cluster A and the C-terminal region.

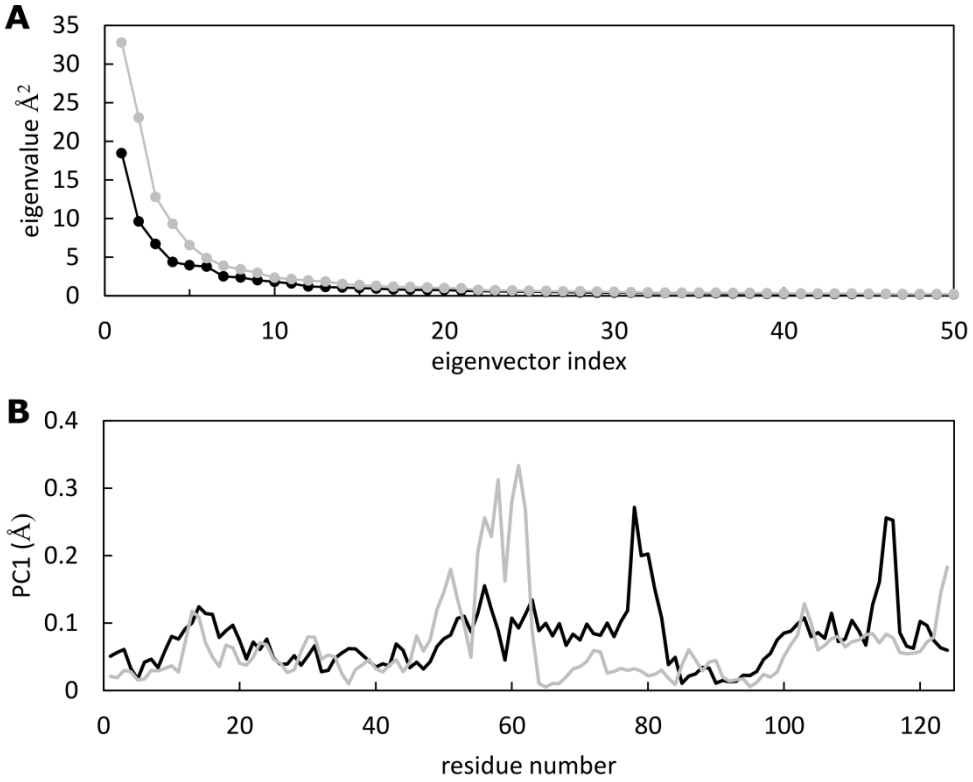

Figure 4. The results of principal component analysis. (A) The eigenvalues plotted against the corresponding eigenvector indices obtained from the Cα covariance matrix constructed from the last 40 ns of the MD trajectory. (B) The displacements of PC1. Black-WT, Grey-R123H.

The dynamic cross-correlation map (DCCM) that reflects the fluctuations of the coordinates of Cα atoms was also calculated to analyze the collective motions of wild type and mutant sPLA2-IIA. Interestingly, the regions 45-52 and 110-124 of the protein appear to be correlated in their motion, as highlighted by the map reported in Figure 5. Moreover, in the case of the mutant (Fig. 5B) the mutation R123H not only induces a stronger (between 0.5 and 0.75) positive correlation between residues 123-124 and 42-52 but also wider anti-correlated motion of the active region 42-52 with most of residues in the C-terminal region (110-124) (Fig. 5B). In comparison with the mutant, motions between these regions were found weaken in wild type sPLA2-IIA (Fig. 5A). This result also suggests that the conformational changes in 3D space of the mutant are higher than the wild type.

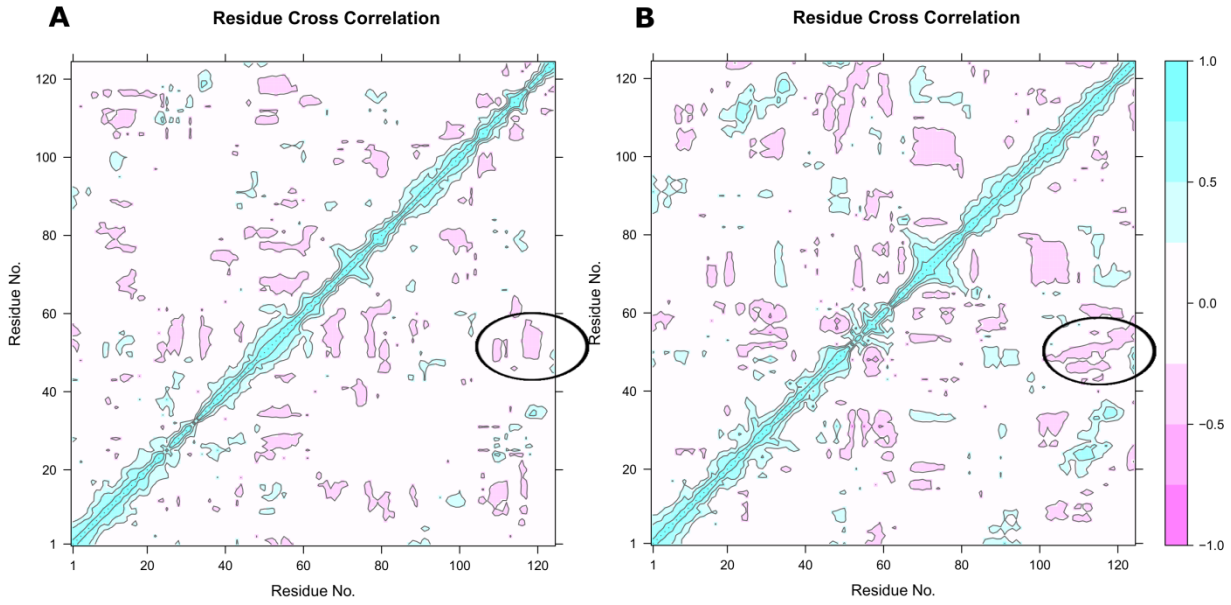

Figure 5. Dynamic cross correlation map of wild type (A) and R123H mutant (B) averaged over the last 40 ns simulations. The extent of the correlation between residues is shown in figure. The color scheme runs between pink (-), white (0) and light blue (+). The negative value indicates anti-

correlation i.e. the C-α atom displacement is in opposite direction, positive value indicates correlated motion i.e. atom displacement in same direction. Deeper color indicates more correlated (or anti-correlated).

3.3.3. Electrostatic properties of sPLA2-IIA and R123H sPLA2-IIA

The presence of a large number of prominent cationic patches on the molecular surface suggests that the electrostatic properties of sPLA2-IIA could be important driving forces in the intermolecular recognition process [41]. We have therefore analyzed the electrostatic potential of sPLA2-IIA and R123H mutant using the program DelPhi based on the coordinates of the average structures over the last 40 ns simulation where the systems display a conformational stability. The analysis outlines a non-uniform electrostatic potential distribution of the enzyme which is characterized by strong positive potential regions due to several clusters of basic amino acids (Fig. 6). Calculations of the electrostatic field for the R123H mutant indicate that the overall distribution of positive potential in the mutant is very similar to that calculated for the WT, except in the region near the side chain of His123 residue where there is a less positive character than in the WT (decrease of 15.96 kT) (Fig. 6A,B). On the side of cluster A (Lys52, Arg53, Lys56 and Arg57) some interesting differences are noted when comparing the two proteins. From the figure it could be acknowledged that in mutant sPLA2-IIA the charged terminal carboxylate of Cys124 is oriented towards Lys52 thus causing a difference in charge distribution on the surface of the protein (Fig 6A, B).

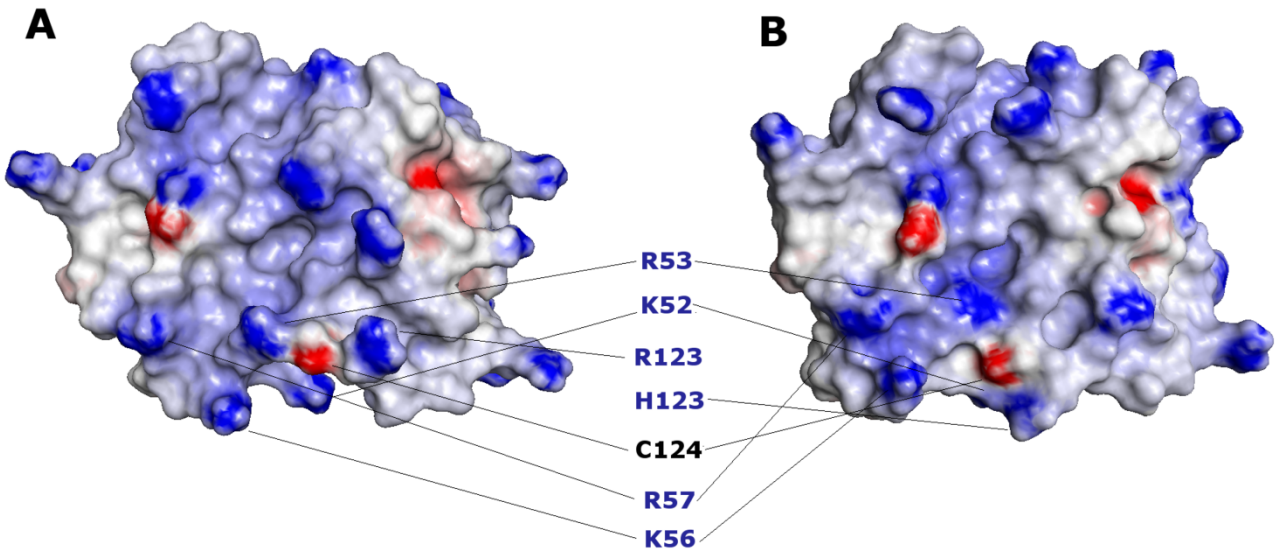

Figure 6. Electrostatic potential. The solvent-accessible molecular surfaces of WT and R123H sPLA2 are shown in a similar orientation. The surfaces are color-coded according to electrostatic surface potential calculated by DelPhi; red, -10 kT/e; white, 0 kT/e; blue +10 kT/e. Cluster A residues are labeled in blue.

We then used *in silico* structure-based pKa prediction to determine the pKa of ionizable amino acid residues in the average conformations. Titrable residues within a radius of 12 Å from the mutated position which showed significant pKa shifts (>0.4) upon R123H mutation as well as the catalytic residues are reported in Table 2. The calculated pKa shifts are a function of the local environment around the mutated residue, depending on the degree of burial, polar and ionic interactions with neighboring residues. The results show that the pKa values for the key catalytic residues His47 and Asp48 were relatively stable suggesting that the local environments around these residues is largely conserved upon mutation (Table 2). Likewise, the change in pKa observed for the catalytic Asp91 and a number of ionizable amino acid residues at a short distance from 123 position (Table 2) reflects the change in the local environment upon mutation. Therefore we further explored the average model structure generated from the MD runs to try to identify the structural mechanisms guiding the pKa shifts.

Table 2

Changes in predicted pKa values

	pKa		ΔpKa
Residue	**Wild-type**	**Mutant**	
His47	7.70	7.44	-0.25
Asp48	3.84	3.85	0.01
Tyr51	12.20	10.30	-1.90
Lys52	10.73	11.15	0.42
Asp91	2.71	2.00	-0.71
Cys124	2.57	2.00	-0.57

3.3.4. Key distances

Since the R123H substitution is about 15 Å from the catalytic His47, it is likely that the active sites of the two sPLA2-IIA are structurally identical. Examining of the average structures of WT and mutant sPLA2-IIA over the last 40 ns simulations (Fig. 7A, B) revealed that the observed Arg/His substitution impacts in terms of catalytic contacts only for the orientation of Tyr51 which is different in WT sPLA2-IIA compared to the mutant one (Fig. 7A,B). The conformation at the active site, upon R123H mutation, is well maintained with the exception of the hydrogen bond pattern of Asp91 which appears modified (Fig. 7A,B). In the wild type protein (both crystallographic and modelled) the carboxylate group of Asp91 is hydrogen bonded to His47, Tyr51 and Tyr66 (Fig. 1 and Fig. 7A). Residues His47, Tyr51, Tyr66 and Asp91 are collectively termed “catalytic network” of PLA2 [42] which, connected by hydrogen bonding network to the N-terminal region of the protein, has been suggested to be involved in the "interfacial recognition site” [42,43]. Upon mutation, Asp91 is hydrogen bonded to His47 and Tyr66 but not to Tyr51 (Fig. 7B). The “catalytic network” is therefore altered, with a probable effect on the interfacial binding properties of PLA2. Average key distances over the entire simulation time have been calculated and shown in Fig. 8. These values are confirmed by the replicate MD simulations (Fig. S4).

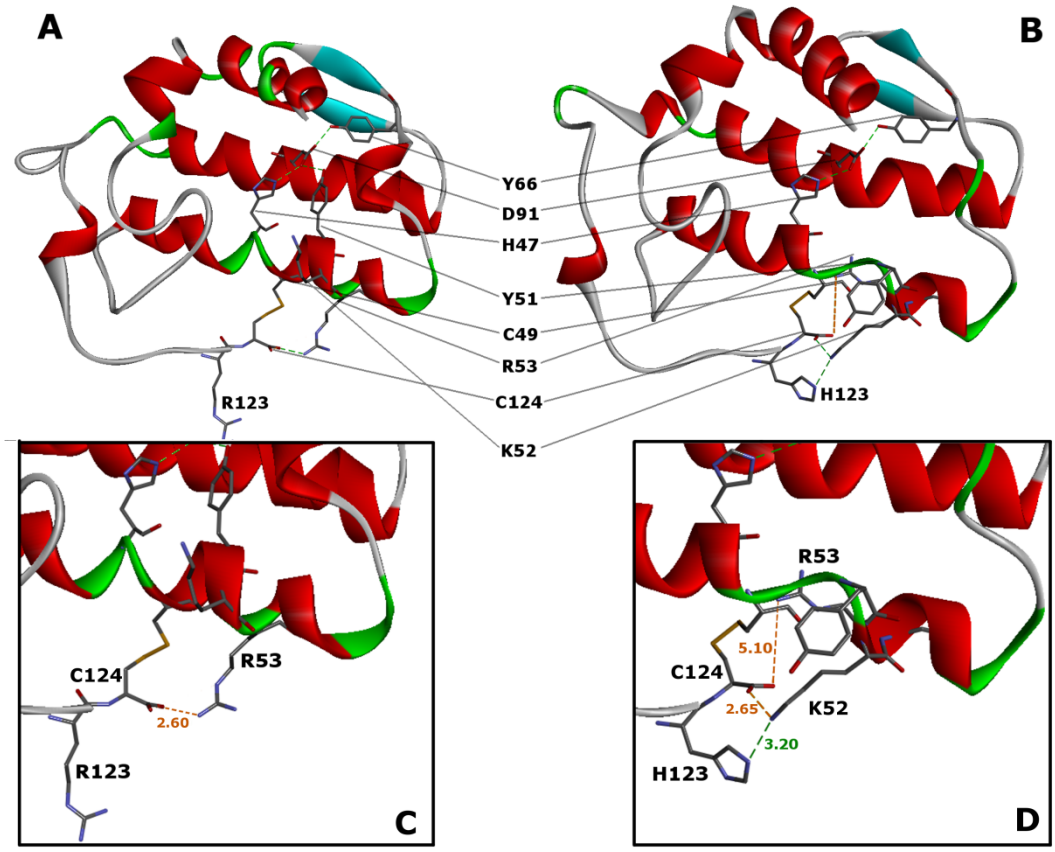

Figure 7. Three-dimensional model structure of WT and R123H mutant sPLA2. The average structure from the last 40 ns of simulation is shown. The protein backbone is represented as solid ribbons colored by secondary structure and the atomic details are represented as stick, colored by atom type. In panel A is reported the structure of the wild-type and in panel B the structure of the mutant. Key residues of the cluster A are highlighted in the enlarged sections C (wild type) and D (mutant). Salt bridges are shown in orange, hydrogen bonds in green.

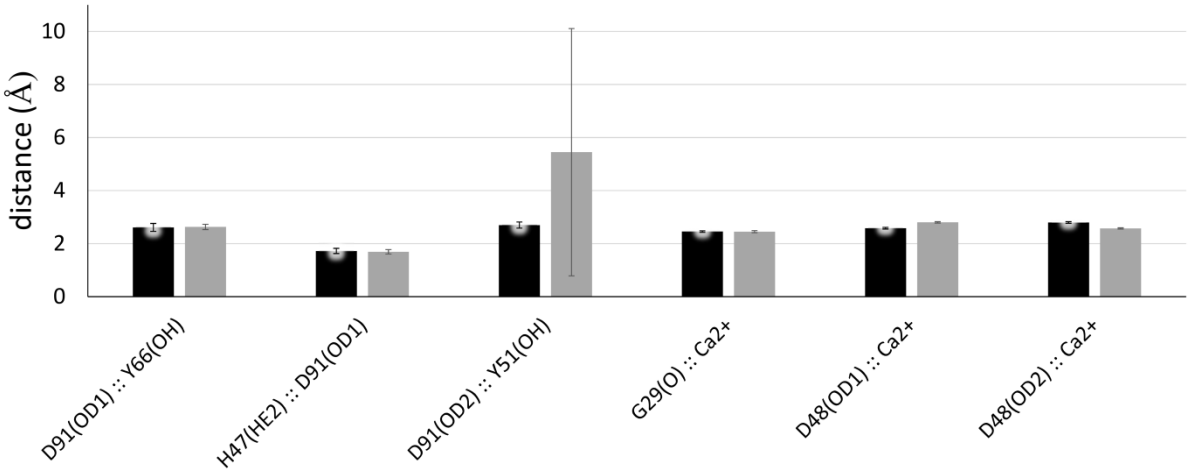

Figure 8. Key distances. Average distances with respective standard deviations were calculated over the entire simulation time. Black-WT, Grey-R123H.

Notably, the OH(Tyr51) to OD2(Asp91) distance is 2.61±0.11 Å (2.60 in the crystal structure) and 5.44±4.58 Å, in the WT and average mutant protein, respectively. Detailed analysis of MD trajectories shows that, over the last 60 ns, Tyr51 breaks its hydrogen bond with Asp91 until it reaches a distance of 14 Å at the end of the simulation (Supplementary Information Fig. S5). Positions of salt bridges and hydrogen bonds and their stability during MD simulations were also checked. Analysis of MD trajectories clearly indicate that in the above mentioned cluster A there are possible and structurally important electrostatic interactions between Cys124(OXT)-Lys52(NZ), Cys124(OXT)-Arg53(NH1) and His123(NE2)-Lys52(NZ). The limit of 6 Å between O···N atoms was accepted for weak salt bridge. It is evident from Figure 7D that the salt bridges Cys124-Lys52 and Cys124-Arg53 of the mutated sPLA2-IIA, were found strong (low distance between donor and acceptor atom) and extremely stable during MD simulation. In contrast, the Cys124(OXT)-Lys52(NZ) distance spends considerable time away from the salt bridge distance during the WT simulation in which, instead, a salt bridge is formed between Cys124 and Arg53 (Fig. 7C). Notably, in the mutant protein, Lys52 also forms an hydrogen bond with His123 and the interaction between the NE2 atom of His123 and the NZ atom of Lys52 is shown in Fig. 7D. The larger number of

electrostatic interactions at this region introduces more bonding constrains that could affect local flexibility. The involvement of Cys124 and His123 in stable interactions with Lys52 and Arg53 and the induced conformational changes at the C-terminus have the effect of steering the backbone of helix 39-56 away from the active site. As a result, a decrease in helix content in the region was observed (Fig. 3) as well as a significant shift of the catalytic Tyr51 which leads to the disruption of the H-bond with Asp91 (Fig. 7B, Fig. 8). Therefore, it may be inferred that changes in mobility in the C-terminal region of the sPLA2-IIA could be propagated into the active site region, which in turn may affect the enzymatic activity of the protein. This result is confirmed by the duplicate MD simulation in which His123 forms electrostatic (pi-anion) interactions with Lys52 (Supplementary Information Fig. S6).

4. Conclusions

We have investigated the consequences of a mutation within the PLA2G2A gene, detected in 2 infants with acute respiratory distress syndrome (ARDS). *In silico* analysis of the novel R123H mutation in sPLA2-IIA indicate that only slight differences in overall 3D structure were observed for R123H mutant compared to wild type. Nevertheless, differences in motions and electrostatic potential distribution at the interfacial binding surface have been observed with possibilities for changes in locations and structure of interaction sites with anionic surfaces including anionic phospholipid membranes, heparinoid, and anionic proteins, which may be important for the physiological functions of the protein. Beyond affecting intermolecular interactions, the observed mutation could lead to an altered catalytic activity due to a different hydrogen bond pattern at the active site. ARDS is a multifactorial condition needing both acute triggers and an underlying terrain, this latter being less elucidated than the possible triggering factors. Our study provides new insights into the genetic predisposition to develop ARDS process, although the proposed mechanism shall require validation through experimental approaches. *In silico* predictions can be used to direct protein engineering and activity experiments.

Competing interests

We declare that we have no competing interests.

Acknowledgements

The authors acknowledge the National Research Council (CNR).

References

1. G. Lambeau and M.H. Gelb, Biochemistry and physiology of mammalian secreted phospholipases A2. Annu. Rev. Biochem. 77 (2008) 495-520.

2. E. Kitsiouli, G. Nakos and M.E. Lekka, Phospholipase A2 subclasses in acute respiratory distress syndrome. Biochim. Biophys. Acta 1792 (2009) 941-953.

3. L. Touqui and L. Arbibe, A role for phospholipase A2 in ARDS pathogenesis. Mol. Med. Today 5 (1999) 244-249.

4. S. Furue, K. Kuwabara, K. Mikawa, K. Nishina, M. Shiga, N. Maekawa, M. Ueno, Y. Chikazawa, T. Ono, Y. Hori, A. Matsukawa, M. Yoshinaga and H. Obara, Crucial role of group IIA phospholipase A(2) in oleic acid-induced acute lung injury in rabbits. Am. J. Respir. Crit. Care Med. 160 (1999) 1292-1302.

5. S. Furue, K. Mikawa, K. Nishina, M. Shiga, M. Ueno, Y. Tomita, K. Kuwabara, I. Teshirogi, T. Ono, Y. Hori, A. Matsukawa, A. Yoshinaga and H. Obara, Therapeutic time-window of a group IIA phospholipase A2 inhibitor in rabbit acute lung injury: correlation with lung surfactant protection. Crit. Care Med. 29 (2001) 719-727.

6. D.K. Kim, T. Fukuda, B.T. Thompson, B. Cockrill, C. Hales and J.V. Bonventre, Bronchoalveolar lavage fluid phospholipase A2 activities are increased in human adult respiratory distress syndrome. Am. J. Physiol. 269 (1995) L109-118.

7. G. Nakos, E. Kitsiouli, E. Hatzidaki, V. Koulouras, L. Touqui and M.E. Lekka, Phospholipases A2 and platelet-activating-factor acetylhydrolase in patients with acute respiratory distress syndrome. Crit. Care. Med. 33 (2005) 772-779.

8. D. De Luca, E. Capoluongo, V. Rigo and Study group on Secretory Phospholipase in P., Secretory phospholipase A2 pathway in various types of lung injury in neonates and infants: a multicentre translational study. BMC Pediatr. 11 (2011) 101.

9. D. De Luca, A. Minucci, P. Cogo, E.D. Capoluongo, G. Conti, D. Pietrini, V.P. Carnielli and M. Piastra, Secretory phospholipase A(2) pathway during pediatric acute respiratory distress syndrome: a preliminary study. Pediatr. Crit. Care Med. 12 (2011) e20-4.

10. D. De Luca, E. Lopez-Rodriguez, A. Minucci, F. Vendittelli, L. Gentile, E. Stival, G. Conti, M. Piastra, M. Antonelli, M. Echaide, J. Perez-Gil and E.D. Capoluongo, Clinical and biological role of secretory phospholipase A2 in acute respiratory distress syndrome infants. Crit. Care 17 (2013) R163.

11. D. De Luca, M. Piastra, G. Chidini, P. Tissieres, E. Calderini, S. Essouri, A. Medina Villanueva, A- Vivanco Allende, M. Pons-Odena, L. Perez-Baena and et al., The use of the Berlin definition for acute respiratory distress syndrome during infancy and early childhood: multicenter evaluation and expert consensus. Intensive Care Med. 39 (2013) 2083-2091.

12. M. Adamzik, U.H. Frey, K. Rieman, S. Sixt, M. Beiderlinden, W. Siffert, and J. Peters, Insertion/deletion polymorphism in the promoter of NFKB1 influences severity but not mortality of acute respiratory distress syndrome. Intensive Care Med. 33 (2007) 1199-1203.

13. E.K. Bajwa, P.C. Cremer, M.N. Gong, R. Zhai, L. Su, B.T. Thompson and D.C. Christiani, An NFKB1 promoter insertion/deletion polymorphism influences risk and outcome in acute respiratory distress syndrome among Caucasians. PLoS One 6 (2011) e19469.

14. N. Takabatake, M. Sata, S. Inoue, Y. Shibata, S. Abe, T. Wada, J. Machiya, G. Ji, T. Matsuura, Y. Takeishi, M. Muramatsu and I. Kubota, A novel polymorphism in secretory phospholipase A2-IID is associated with body weight loss in chronic obstructive pulmonary disease. Am. J. Respir. Crit. Care. Med. 172 (2005) 1097-1104.

15. R.W. Schevitz, N.J. Bach, D.G. Carlson, N.Y. Chirgadze, D.K. Clawson, R.D. Dillard, S.E. Draheim, L.W. Hartley, N.D. Jones, E.D. Mihelich and et al., Structure-based design of the first potent and selective inhibitor of human non-pancreatic secretory phospholipase A2. Nat. Struct. Biol. 2 (1995) 458-465.

16. Y. Snitko, R.S. Koduri, S.K. Han, R. Othman, S.F. Baker, B.J. Molini, D.C. Wilton, M.H. Gelb and W. Cho, Mapping the interfacial binding surface of human secretory group IIa phospholipase A2. Biochemistry 36 (1997) 14325-14333.

17. D.L. Scott, S.P. White, J.L. Browning, J.J. Rosa, M.H. Gelb and P.B. Sigler, Structures of free and inhibited human secretory phospholipase A2 from inflammatory exudate. Science 254 (1991) 1007-1010.

18. J.P. Wery, R.W. Schevitz, D.K Clawson, J.L. Bobbitt, E.R. Dow, G. Gamboa, T Goodson Jr, R. B Hermann, R. M Kramer, D. B McClure and et al., Structure of recombinant human rheumatoid arthritic synovial fluid phospholipase A2 at 2.2 Å resolution. Nature 352 (1991) 79–82.

19. R.S. Koduri, S.F. Baker, Y. Snitko, S.K. Han, W. Cho, D.C. Wilton and M.H. Gelb, Action of human group IIa secreted phospholipase A2 on cell membranes. Vesicle but not heparinoid binding determines rate of fatty acid release by exogenously added enzyme. J. Biol. Chem. 273 (1998) 32142-32153.

20. J.M. Winget, Y.H. Pan and B.J. Bahnson, The interfacial binding surface of phospholipase A2s. Biochim. Biophys. Acta 1761 (2006) 1260-1269.

21. S. Piana, A. Laio, F. Marinelli, M. Van Troys, D. Bourry, C. Ampe and J C. Martins, Predicting the effect of a point mutation on a protein fold: the villin and advillin headpieces and their Pro62Ala mutants. J. Mol. Biol. 375 (2008) 460-470.

22. S. Lindert, Y. Cheng, P. Kekenes-Huskey, M. Regnier and J. A. McCammon, Effects of HCM cTnI mutation R145G on troponin structure and modulation by PKA phosphorylation elucidated by molecular dynamics simulations. Biophys. J. 108 (2015) 395-407.

23. D. Pirolli, F. Sciandra, M. Bozzi, B. Giardina, A. Brancaccio and M. C. De Rosa, Insights from molecular dynamics simulations: structural basis for the V567D mutation-induced instability of zebrafish alpha-dystroglycan and comparison with the murine model. PLoS ONE 9 (2014) e103866.

24. D. Pirolli, C. Carelli Alinovi, E. Capoluongo, M A. Satta, P. Concolino, B. Giardina and M. C. De Rosa 4, Insight into a novel p53 single point mutation (G389E) by molecular dynamics simulations. Int. J. Mol. Sci. 12 (2011) 128-140.

25. G.R. Bernard, A. Artigas, K.L. Brigham, J. Carlet, K. Falke, L. Hudson, M. Lamy, J.R. Legall, A. Morris and R. Spragg, The American-European Consensus Conference on ARDS. Definitions, mechanisms, relevant outcomes, and clinical trial coordination. Am. J. Respir. Crit. Care Med. 149 (1994) 818-824.

26. R.M. Kramer, C. Hession, B. Johansen, G. Hayes, P. McGray, E.P. Chow, R. Tizard and R.B. Pepinsky, Structure and properties of a human non-pancreatic phospholipase A2. J. Biol. Chem. 264 (1989) 5768-5775.

27. I.A. Adzhubei, S. Schmidt, L. Peshkin, V.E. Ramensky, A. Gerasimova, P. Bork, A.S. Kondrashov, and S.R. Sunyaev, A method and server for predicting damaging missense mutations. Nat. Methods 7 (2010) 248-249.

28. P.C. Ng and S. Henikoff, SIFT: Predicting amino acid changes that affect protein function. Nucleic Acids. Res. 31 (2003) 3812-3814.

29. R.A. Laskowski, D.S. Moss and J.M. Thornton, Main-chain bond lengths and bond angles in protein structures. J. Mol. Biol. 231 (1993) 1049-1067.

30. E. Feyfant, A. Sali and A. Fiser, Modeling mutations in protein structures. Protein Sci. 16 (2007) 2030-2041.

31. R. Luthy, J.U. Bowie and D. Eisenberg, Assessment of protein models with three-dimensional profiles. Nature 356 (1992) 83-85.

32. M. Wiederstein and M.J. Sippl, ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic Acids. Res. 35 (2007) W407-410.

33. D. Shivakumar, J. Williams, Y. Wu, W. Damm, J. Shelley and W. Sherman, Prediction of absolute solvation free energies using molecular dynamics free energy perturbation and the OPLS force field. J. Chem. Theory Comput. 6 (2010) 1509-1519.

34. W.G. Hoover, Canonical dynamics: Equilibrium phase-space distributions. Phys. Rev. A. Gen. Phys. 31 (1985) 1695-1697.

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

36. B. Grant, A.P.C. Rodrigues, K.M. ElSawy, J.A. McCammon, L.S.D. Caves, Bio3D: an R package for the comparative analysis of protein structures. Bioinformatics 22 (2006) 2695-2696.

37. L. Li, C. Li, S. Sarkar, J. Zhang, S. Witham, Z. Zhang, L. Wang, N. Smith, M. Petukh and E. Alexov, DelPhi: a comprehensive suite for DelPhi software and associated resources. BMC Biophys. 5 (2012) 9.

38. V.Z. Spassov and L. Yan, A fast and accurate computational approach to protein ionization. Protein. Sci. 17 (2008) 1955-1970.

39. A. Abu-Maziad, K. Schaa, E.F. Bell, J.M. Dagle, M. Cooper, M.L. Marazita and J.C. Murray, Role of polymorphic variants as genetic modulators of infection in neonatal sepsis. Pediatr. Res. 68 (2010) 323-329.

40. E. Movert, Y. Wu, G. Lambeau, F. Kahn, L. Touqui and T. Areschoug, Secreted group IIA phospholipase A2 protects humans against the group B streptococcus: experimental and clinical evidence. J. Infect. Dis. 208 (2013) 2025-35.

41. C.N. Birts, C.H. Barton, D.C. Wilton DC. Catalytic and non-catalytic functions of human IIA phospholipase A2. Trends Biochem Sci. 2010 Jan;35(1):28-35.

42. C.M. Dupureur, T.L. Deng, J.G. Kwak, J.P. Noel and M.D. Tsai, Phospholipase-A2 engineering. 4. Can the active-site aspartate-99 function alone ? J. Am. Chem. Soc. 112 (1990) 7074-7076.

43. D.L. Scott, S.P. White, Z. Otwinowski, W. Yuan, M.H. Gelb and P.B. Sigler, Interfacial catalysis: the mechanism of phospholipase A2. Science 250 (1990) 1541-1546.