Molecular Dynamics simulation of the heart type Fatty Acid Binding Protein in a crystal environment

Yanis R. Espinosa, H. Ariel Alvarez, Eduardo I. Howard & C. Manuel Carlevaro

Accepted author version posted online: 25 May 2020.

Submit your article to this journal

View related articles

View Crossmark data

Research Article

Molecular Dynamics simulation of the heart type Fatty Acid Binding Protein in a crystal environment

Yanis R. Espinosa[a,b], H. Ariel Alvarez[a,c,d], Eduardo I. Howard[a,e,*] and C. Manuel Carlevaro[a,f,*]

[a]Instituto de Física de Líquidos y Sistemas Biológicos (CONICET-UNLP), Calle 59 Nro 789, B1900BTE La Plata, Argentina.

[b]Grupo de Bioquímica Teórica, Universidad Industrial de Santander, Cra 27, Calle 9, Bucaramanga, Colombia.

[c]Departamento de Ciencias Biológicas, Facultad de Ciencias Exactas, UNLP, Calle 47 y 115, 1900 La Plata, Argentina.

[d]Instituto de Ciencias de la Salud, Universidad Nacional Arturo Jauretche, Av. Calchaquí 6200, 1888 Florencio Varela, Buenos Aires, Argentina.

[e]Universidad Tecnológica Nacional- Facultad Regional Tierra del Fuego, Perito Francisco Moreno 1415, V9410KSY Ushuaia, Tierra del Fuego, Argentina

[f]Departamento de Ingeniería Mecánica, Universidad Tecnológica Nacional, Facultad Regional La Plata, Av. 60 Esq. 124, La Plata, 1900, Argentina

*Corresponding authors Eduardo I. Howard howard@iflysib.unlp.edu.ar

Contact: C. M. Carlevaro. Instituto de Física de Líquidos y Sistemas Biológicos (CONICET-UNLP), Calle 59 Nro 789, B1900BTE La Plata, Argentina. Email: manuel@iflysib.unlp.edu.ar

Abstract

Crystallographic data comes from a space-time average over all the unit cells within the crystal, so dynamic phenomena do not contribute significantly to the diffraction data. Many efforts have been made to reconstitute the movement of the macromolecules and explore the microstates that the confined proteins can adopt in the crystalline network. We explored different strategies to simulate a heart fatty acid binding protein (H-FABP) crystal by means of Molecular Dynamics (MD) simulations. We evaluate the effect of introducing restraints according to experimental isotropic B-factors and we analyzed the H-FABP motions in the crystal using Principal Component Analysis (PCA), isotropic and anisotropic B-factors. We compared the behavior of the protein simulated in the crystal confinement versus in solution, and we observed the effect of that confinement in the mobility of the protein residues. Restraining one-third of $C\alpha$ atoms based on experimental B-factors produce lower B-factors than simulations without restraints, showing that the position restraint of the

atoms with the lowest experimental B-factor is a good strategy to maintain the geometry of the crystal with an obvious decrease in the degrees of motion of the protein. PCA shows that, as position restraint reduces the conformational space explored by the system, the motion of the crystal is better recovered, for an essential subspace of the same size, in the simulations without restraints. Restraining only one *Cα* seems to be a good balance between giving flexibility to the system and preserving its structure.

Abbreviations:

H-FABP, heart fatty acid binding protein; MD, Molecular Dynamics; PDB, Protein Data Bank; POPC, 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine; PCA, principal component analysis; NVT, constant Number of atoms, Volume and Temperature; NpT, constant Number of atoms, pressure and Temperature.

1 Introduction

X–ray crystallography has been the major contributor to our knowledge of the structure of macromolecules (Berman et al., 2013). At the moment, almost 90% of the structures deposited in the Protein Data Bank (Berman et al., 2000) (PDB) have been solved by this technique, which has conditioned our way of representing macromolecules, offering its vision, as well as its limitations. The crystallographic data comes from a space-time average over all the unit cells in the crystal, so that the dynamic phenomena in an individual unit cell does not contribute significantly to the diffraction data, which are finally interpreted in terms of a mean structure (Kruschel and Zagrovic, 2009).

This single model representation is further reinforced by the fact that the crystal lattice hinders diffusion and restricts macromolecular fluctuations (Andrec et al., 2007). Many efforts have been made to *reconstitute the movement* of the macromolecules and explore the microstates that the confined proteins can adopt in the crystalline network (Janowski et al., 2013; Kuzmanic and Zagrovic, 2010; Ma et al., 2015). Experimental approaches and different modeling techniques have been developed to recover this information (Kuzmanic et al., 2014; Li et al., 2014b; Wall

et al., 2014; Xue and Skrynnikov, 2014). Among the computational tools, the use of Molecular Dynamics (MD) simulations and Normal Mode Analysis (NMA) (Cerutti et al., 2010; Kondrashov et al., 2007; Meinhold and Smith, 2005; Terada and Kidera, 2012) has introduced major advances. MD simulations and NMA have the potential to provide information on dynamics and heterogeneity hidden in the X–ray diffraction data (Janowski et al., 2013). Moreover, Normal Mode Analysis offers an efficient way to study the conformational flexibility of protein structures (Kondrashov et al., 2007).

An interesting case is the family of the Fatty Acid Binding Proteins (FABPs), whose activity requires that conformational flexibility. The proteins of this family are involved in the traffic of fatty acids inside the cell and, despite extensive studies, the entry/release mechanism of the transported fatty acids is not well understood. Most of the structures in the Protein Data Bank show the fatty acid inside the cavity of the protein. This is compatible with the transport of lipids, but also reveals that a conformational change is necessary to open a portal to the outside (See Figure 1).

The high-resolution structure of H-FABP was recently obtained at room temperature using the combined techniques of X–ray and neutron diffraction (Howard et al., 2016), placing this structure in a privileged position in terms of diffraction quality and techniques used. The experimental results obtained are highly reliable, and even so the analysis of the Atomic Displacement Parameters (ADPs) does not reveal the opening mechanism of the portal. Also, the structures obtained by NMR enrich our knowledge of this family of proteins, but they do not give an obvious answer about the problem (Cai et al., 2012).

In this context, it is not clear whether the lack of information about the intermediate states necessary for the activity in the structures deposited in the PDB is the result of crystalline packing, that is, intrinsic to the limitations to movement imposed by the confinement of the protein within the crystal, or it is the result of diffraction techniques that give us an average structure that has

discarded all details of microstates. Molecular Dynamics simulation could be a suitable tool to address this problem.

The study of proteins in solution by means of MD simulations is a well-documented task, and in particular FABP has been the subject of several studies (Chen et al., 2014; Friedman et al., 2006; Levin et al., 2010; Likić and Prendergast, 1999; Long et al., 2009; Tsfadia et al., 2007; Yan et al., 2018a,b; Zamarreño et al., 2018, 2012). But simulating a protein crystal is not so well documented. There are few examples in the literature and several variables that can be adjusted. In the present work we explore different simulation strategies to study a protein crystal by MD simulations (i.e., NVT or NpT ensembles, with or without position restraints), and we compare the results with the simulation of the protein in solution, giving a possible answer to the problem of the intermediate states mentioned above.

2 Computational methods

2.1 Molecular Dynamics

When the structure of a macromolecule is solved by diffraction techniques, the positions of the atoms that have been identified in the asymmetric unit are deposited in the PDB, along with the information about crystallographic space group and its related symmetries. To model a crystal, it is necessary to use these symmetries to reconstruct the content of the unit cell and then, applying periodic boundary conditions, we are able to simulate an *infinite*, borderless crystal.

We have generated the initial coordinates of our crystal, starting from the PDB ID 5CE4, an X-ray/neutron diffraction structure collected at room temperature. Using PyMOL (DeLano and Bromberg, 2004) (*symexp* command) we have applied the symmetry operations of the $P2_12_12_1$ space group to the protein and all structural water molecules identified (crystallographic waters). The values for the unit cell dimensions were a = 3.4588 nm, b = 5.5307 nm, c = 7.1185 nm. Considering that the length of the X axis is close to the cut-off used during the simulation (*i.e.*, 1.2 nm), we doubled the cell in this direction

to avoid self-influence across periodic boundary conditions, and by this way the initial box dimensions were 2 × 1 × 1 of $2L_X$ = 6.9176 nm, L_Y = 5.5307 nm, L_Z = 7.1185 nm. Hence, the simulation box contained eight H-FABP molecules, each complexed with a fatty acid (four complexes per unit cell) and 3972 SPC/E water molecules (Berendsen et al., 1987), from which 1376 were crystallographic water molecules. Fatty acid content in H-FABP depends on its cellular abundance in the organism in which the protein is expressed, and the crystallographic results are not clear regarding the identity of the fatty acid, due to the flexibility of the last part of the tail. Due to these facts, four of the eight simulated proteins contain palmitic acid, and the other four contain oleic acid (*i.e.*, 4 H-FABP–palmitic acid complexes and 4 H-FABP–oleic acid complexes in the simulation box).

The effective pH was assumed to be 7.5, same as in the crystallization buffer. The protonation status of individual Asp, Glu, Lys, Arg, and His residues was obtained by PROPKA (Olsson et al., 2011) calculations for H-FABP in a crystal-lattice environment, leading to a charge of –1 per H-FABP molecule. Thus, the net charge of each H-FABP–fatty acid complex was –2, so sixteen Na$^+$ counterions were added to neutralize the total charge of the system. The system was simulated using the united-atom GROMOS 54A7 force field (Schmid et al., 2011). Parameters for topologies of palmitic and oleic acid were obtained from Tsfadia and cols. (2007)(Tsfadia et al., 2007) and from POPC (1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine) parametrization for this force field, and added to it (see Supplemental Figure 1 and topologies incorporated as Supplementary material for details on their parametrization).

The energy of the simulated system was initially minimized following a process where we applied 500 steps of steepest descent algorithm until a potential energy gradient ΔE ≤ 1000 kJ mol^{-1} was achieved. The protein and lipid atoms being harmonically restrained to their initial positions with a force constant of 25,000 kJ mol^{-1} nm^{-2} in all Cartesian directions. After assigning random initial velocities from a Maxwell-Boltzmann distribution at 100 K, the system was subsequently heated in three steps of 50 K and one step of 43 K, up to 293 K, simulating during 100 ps for each step. Simultaneously, for the

same time lapse, the atomic position restraints in each protein molecule were uniformly relaxed down to zero (harmonic potential force constant relaxed from 25,000 to 0 kJ mol^{-1} nm^{-2} in steps of 5,000 kJ mol^{-1} nm^{-2}). The Cα atoms from residues with a temperature factor (B–factor) lower than a value near 10 (44 atoms, one third of the total Cα atoms) were kept restrained (NVT with 44 Cα atoms restraint) throughout these equilibration runs using a restraining elastic constant of 25,000 kJ mol^{-1} nm^{-2} (see Table 1 in Supplementary material for details of B–factor values for each atom). The equilibration runs were performed at constant volume.

Four different schemes at 293 K were applied for the treatment of the crystal unit cell volume and the deformations on the lattice:

- NpT without restraints,
- NVT without restraints,
- NVT with 44 Cα atoms restraint, and
- NVT with 1 Cα atom restraint.

The Cα atom selected to restraint in the last scheme is the one corresponding to residue ILE114. This residue has a low isotropic B–Factor, almost symmetric anisotropic B–Factors and is far from the residues that we considered in our analysis. Finally, with the aim of analyzing the effect of crystallographic packing on the mobility of the residues, we simulated a single protein solution using an NpT ensemble without restrictions (same temperature and pressure as in the crystal simulations), following the minimization and equilibration protocol described in this section. All production simulations were run for 500 ns for each scheme using the GROMACS 2016.3 biomolecular simulation package (Abraham et al., 2015) with a 2 fs integration time step. During equilibration and production, protein and non-protein groups were coupled separately to a heat bath using the Velocity–rescale thermostat (Bussi et al., 2007) with a relaxation time of 0.05 ps. In the NpT ensemble simulations, the pressure was coupled to a Parrinello–Rahman barostat (Parrinello and Rahman, 1981) at 1 bar with a relaxation time of 1.0 ps. The bond lengths were constrained using LINCS algorithm (Hess

et al., 1997) while electrostatic interactions were computed using the Particle Mesh Ewald method (Abraham and Gready, 2011). A cut-off of 1.2 nm was applied both for the van der Waals and Coulomb interactions with a Verlet cut-off scheme. All calculations were carried out on a Linux server Intel Core i7-6700 3.40 GHz eight Core Processor with a NVIDIA GeForce GTX 1080 GPU.

2.2 Essential dynamics: Principal component analysis

Principal component analysis (PCA) reveals the most important motions in proteins. Thus, a collective coordinates set is obtained from the atomic fluctuations on the protein, commonly used to predict a low-dimensional subspace in which essential protein motion takes place (Daidone and Amadei, 2012). As of a linear transform that describes the accessible degrees of freedom in the protein, the PCA constructs, from fluctuations of main-chain atoms, a covariance matrix that describes collective modes of fluctuation of the positions of these ones in the protein (from this covariance matrix, eigenvectors and eigenvalues can be computed)(Amadei et al., 1993). Sorting the eigenvectors by the size of their eigenvalues, shows that the configurational space can be divided in a low dimensional *essential* subspace in which most of the positional fluctuations are confined (Daidone et al., 2005).

Therefore, the correlation between each element of the covariance matrix C can be represented as (Amadei et al., 1993; Maisuradze et al., 2010):

$$C_{i,j} = \langle x_i - \langle x_i \rangle \rangle \langle x_j - \langle x_j \rangle \rangle, \quad (1)$$

where x_i and x_j are the mass-weighted Cartesian coordinates of an N-particle system ($x_1,\ldots,x_{3N}$) and $\langle \rangle$ represents the average over all instantaneous structures sampled during simulation time. Thus, the symmetric $3N \times 3N$ matrix C can be diagonalized with an orthonormal transformation matrix T:

$$q = T^T (x - \langle x \rangle), \quad (2)$$

which transforms C into a diagonal matrix $\Lambda = \langle q q^T \rangle$ of eigenvalues λ_i:

$$\Lambda = T^T C T = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_{3N}) \qquad (3)$$

where $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_{3N}$. The ith column of T is the eigenvector belonging to λ_i. Thus, the MD trajectory (*i.e.*, the trajectory of each of the atoms in the simulated system) can be projected on the eigenvectors to determine the principal components (PC) $q_i(t), i = 1, \ldots, 3N$. In this way, the first few PCs typically describe collective global motions of the system.

Finally, we can define a new orthonormal basis with the eigenvectors from linear combinations of the original N dynamical observables to describe the states space of the system. Hence, our covariance matrix was calculated using the Cα carbons from the H-FABP crystal during the total time of the trajectory for each scheme simulated.

2.3 B–Factors Calculation

In order to further analyze the behavior of the crystal simulation, we performed the theoretical calculation of the isotropic and anisotropic B–factors (*i.e.*, the mean-square displacements of the atoms, also termed anisotropic displacement parameters - ADPs) for the simulation runs, so as to compare them with their experimental values. They can be obtained from the Root Mean Square Fluctuations (RMSF) of the positions of the atoms during simulations. The average Cα B–factor values were calculated as Kusmanic *et. al.* (2010)(Kuzmanic and Zagrovic, 2010):

$$B = \frac{8\pi^2}{3} \text{RMSF}^2 \qquad (4)$$

Usually, a threshold is defined to separate high from low B–factor values. Following Parthasarathy and Murthy (Parthasarathy and Murthy, 2000), this threshold was obtained from $\langle B \rangle + \alpha.\sigma(B)$, where $\langle B \rangle$ is the mean B–factor of all the $C\alpha$ atoms used in the entry, α is a factor (usually 0.5 or 0.75) and $\sigma(B)$ is the standard deviation of the distribution of B–factor values. Calculated B–factor values for each simulation were normalized as in previous works (K. Balendiran et al., 2014; Parthasarathy and Murthy, 2000):

$$B_{norm} = \frac{B - \langle B \rangle}{\sigma(B)} \qquad (5)$$

With this normalization, as $\langle B_{norm} \rangle$ = 0 and $\sigma(B_{norm})$ =1, the threshold to separate high from low B–factor values becomes equal to α.

The ADPs define the 3 × 3 symmetric atomic mean-square displacement tensor U_{ij}. The isotropic displacement parameter can be computed by $B_{eq} = 8\pi^2 U_{eq} = \frac{8}{3}\pi^2(U_{11} + U_{22} + U_{33})$. As U_{ij} are tensors, the comparison of their experimental with simulated values is more complex than with the isotropic ones, so the six independent elements of the symmetric tensor can be compared in different ways, as described by Yang and cols(Yang et al., 2009). Let U_{ij} and V_{ij} be the two tensors to compare, a clear way to do so is to compute the normalized correlation coefficient $ncc(U_{ij}, V_{ij})$, defined as:

$$ncc(U_{ij}, V_{ij}) = \frac{cc[U, (U_{eq}/V_{eq})V]}{cc[U, U^{iso}]cc[V, V^{iso}]} \quad (6)$$

where $cc[U_{ij}, V_{ij}] = \frac{(\det(U^{-1})\det(V^{-1}))^{\frac{1}{4}}}{\sqrt{(\frac{1}{8})\det(U^{-1} + V^{-1})}}$, U^{iso} and V^{iso} are diagonal matrices that describe a pair of isotropic atoms, with $U_{11}^{iso} = U_{22}^{iso} = U_{33}^{iso} = U_{eq} = Tr(U_{ij})/3$ and similarly for V^{iso} and V_{eq}.

The normalized correlation coefficient *ncc* will have the following values:

- *ncc* > 1 if two atoms described by *U* and *V* are more similar to each other than to an isotropic atom.
- *ncc* ≤1 otherwise.

With *ncc*, we can compare the size, orientation, and direction of two tensors. A detailed description of the geometrical interpretation of *ncc* can be found in reference (Merritt, 1999). If we calculate the ratio of how many atoms in a structure have their *ncc* values larger than 1 and the total number of atoms,

and express it as a percentage, we can give a good measure of the quality of an anisotropic B–factor prediction.

3 Results and discussion

MD simulations were performed using a solvated unit cell model of crystalline H-FABP consisting of two unit cells in a $2 \times 1 \times 1$ layout (See Figure 2). Analyzed trajectories were obtained during 500 ns of production for the ensembles NVT with restraints, NVT and NpT without restraints (See Computational methods).

In all our analysis, we applied both a rotational and a translational fit over the $C\alpha$ atoms into all eight protein molecules of each system (and over the $C\alpha$ atoms of the single protein in solution) in order to reduce the overestimation of the positional fluctuations in the residues (Stocker and van Gunsteren, 2006).

Initially, we analyze the stability of the system calculating the root mean square deviation (RMSD) of the protein atomic positions and root mean square fluctuation (RMSF) of the positions of the $C\alpha$ atoms in each H-FABP residue. In Figure 3A, we show the RMSD values for all the simulation runs. Predictably, the NVT crystal with positions restrained in forty–four of its $C\alpha$ atoms (one third of the total $C\alpha$ atoms) shows the lowest RMSD value (~0.17 nm), while the NVT and NpT systems without position restraint converge quickly with no difference between their RMSD values (~0.27 nm). The NVT crystal with one $C\alpha$ atom per protein restrained converges to an RMSD value slightly smaller than NVT and NpT crystal simulations. The single protein in solution system has more fluctuating RMSD values and has equilibration values slightly bigger than the crystal systems without position restraints(~0.3 nm).

Likewise, in the RMSF values shown in Figure 3B, we observe that in protein crystals at different conditions the movement throughout the systems tends to have similar dynamics, and despite the restraint in the $C\alpha$ atoms, the crystals show a qualitative correlation in their motions, indicating that the position restraint of the atoms with the lowest B–factor is a good strategy to maintain

the geometry of the crystal without losing the relevant motions in the proteins. All the systems without restraints have similar RMSF values, except for the protein in solution system in the zone delimited by residues 15 to 40, for which RMSF values were significantly bigger.

3.1 Essential motions

To better understand the important protein movements occurred in the simulations, we analyzed the trajectories of the $C\alpha$ atoms from H-FABP crystal using principal component analysis (PCA). Thus, it is possible to detail the direction and amplitude of movements which are relevant for the functioning of the proteins (Amadei et al., 1993).

The $C\alpha$ covariance matrices for the eight H-FABP molecules into the crystal were diagonalized to obtain the eigenvectors and their associated eigenvalues. Subsequently, the trajectory for each system was projected onto the eigenvectors to obtain the principal components.

In our analysis, we observed that the top ten components with largest amplitudes, represent 55.09% of the movements for NpT, 51.69% for NVT, 38.57% for NVT with position restraints in $44C\alpha$, 61.31% for NVT with position restraints in $1C\alpha$, and finally, 70.99% for protein in solution. Interestingly, for NVT ensemble with position restraints, the top components with largest amplitudes represent the lowest percentage of movements in crystal even when compared with the first one hundred components from the other systems (See Supplemental Figure 2 in Supplementary material). In this particular case, the position restraints minimize the mobility of atoms, as shown in the Figure 4, at the same time they reduce the fluctuations in the unrestricted atoms in the protein, *i.e.*, the total atomic fluctuation in the crystal is restricted (See Supplemental Figure 3 in Supplementary material).

Moreover, in the visual inspection of Figure 4, the H-FABP molecules without restraints show a cooperative movement, which is mitigated when the atomic movement in the crystal is restricted. Thus, to analyze the degree of stability of the crystal in the conformational space during the simulation, the local

flexibility of each H-FABP molecule was analyzed by calculating the per-residue B–factors (Cα B–factor), before being averaged over the H-FABPs both unit cells, to be subsequently compared with the crystallographic B–factor (See Figure 5).

The MD simulation B–factors analysis (Figure 5-A) showed greater local flexibility. However, although there is an overestimation of the calculated B–factors, except for NVT with position restraints, a good qualitative correlation between the landscapes of the simulated and the experimental B–factors is evident. Quantitatively, in NVT simulation with position restraints, B–factors look significantly underestimated, except in the region around residue 75, where the computational estimates seem to deviate much more from the experimental ones, as in the rest of the simulated conditions. This situation is due to the perturbations caused by fatty acid movements in the cavity, where the tail region interacts with the β-loop (T74, A75, D76, D77, R78) which is denoted in the literature as a portal that provides the structural basis for fatty acids interchange (Li et al., 2014a; Long et al., 2009; Xiao et al., 2016) (see Supplemental Figure 4 in Supplementary material). The good agreement of the normalized B–factor values for the crystal MD simulations can be seen in Figure 5-B. We analyzed the regions with high B–factor values, considering a threshold $\alpha = 0.75$. The residues with normalized B–factor values over the threshold are shown in Figure 5-B (violet region). The residues with normalized B–factors bigger than 0.75 for the NVT ($1C\alpha$) simulation are coloured in violet in Figure 5-C (*i.e.*, residues 22-25, 57, 74-78, 98-101, 110, 111, and 119-122).

In addition, we can analyze the sampling convergence computing the root mean square inner product (RMSIP) as a measure of similarity between subspaces of each system (Papaleo et al., 2009). Thus, the overlap (O) between a given PC vector Y and another PC vector X is evaluated by their normalized projection (Batista et al., 2011),

$$O = \frac{Y \cdot X}{\| Y \| \| X \|}, \qquad (7)$$

where Y and X are PC vectors from two trajectories at different ensembles.

The essential subspace of each system was defined by the first one hundred eigenvectors with higher eigenvalues, which represented 3.31% of the total configurational spaces ($3N = 3192$), recovered around 82.04% (NpT), 81.86% (NVT), 83.28% (NVT with restraints in $1C\alpha$) and 74.34% (NVT with restraints in $44C\alpha$) of the total motions in the crystal. Thus, the overlap between the essential subspace of two different groups was obtained from the RMSIP as,

$$\text{RMSIP} = \frac{1}{100} \left(\sum_{i=1}^{100} \sum_{j=1}^{100} (n_i . v_j) \right)^{1/2}, \qquad (8)$$

where n_i and v_j are the eigenvectors of the subspaces to be compared. RMSIP ranges from 0 to 1. A perfect match of the sampled subspaces yields an overlap value of 1.

According to our analysis, we observed that independently of the ensemble simulated, the RMSIP values were around 0.63–0.76, indicating global patterns of correlated movements and a satisfactory overlap between essential subspaces of each system (Amadei et al., 1999). Moreover, the similarity of essential subspaces tends to be the lowest (between 0.63–0.66) when the systems NpT, and NVT are overlapped with the NVT with position-restraints system (See Table 1), because the H-FABP crystal with position restraints explores a smaller conformational space.

3.2 Displacement Parameters

As seen in Figure 6, when the H-FABP is subjected to the crystallographic packing the fluctuation in its movements is reduced in relation to the H-FABP in solution, observing in addition, fluctuations that differ between regions of the proteins (See Figure 6 C and D). Moreover, as observed in Figure 5, the experimental B–factor is smaller in relation to the simulated ones, keeping a greater similarity with the global movements observed in simulated H-FABP in a crystallographic packing.

The normalized correlation coefficients *ncc* are calculated to compare the experimental anisotropic temperature factors with those predicted by our MD simulations, in order to get a clear picture of the quality of the MD trajectories obtained, that intend to represent a true crystal system. In addition, for comparison purposes, the *ncc* values for the simulated single protein in solution are also shown. In Figure 7 we can see *ncc* values for all simulated conditions (the values for the crystal structures are calculated aligning each protein in the crystal with the experimental structure taken as reference, an then the anisotropic temperature factors are calculated as a temporal mean value along the trajectory). In this figure it can be seen that the *ncc* values with bigger standard deviations correspond to β-loops, in which correlated movements are infrequent (*e.g.*, residues 95 to 100). For all crystal simulations, the percentage of residues with *ncc* > 1 (which means that the prediction is in good agreement with experiment) is high, as can be seen in Table 2. This is not the case for the simulation of the single protein in solution, for which this percentage is considerably lower. Also, in the *per-residue* representation of *ncc* values (Figure 7) it can be seen that many regions in this simulation do not have correlated movements with the experiment (*i.e.*, they have *ncc* values bellow 1). These results show that there exists high similarity between the experimental anisotropic B-factors and the ones predicted by our MD simulations, and also a clear difference between crystal and solution MD simulations.

4 Conclusion and perspectives

In the present work, we have explored different strategies to simulate a protein crystal starting from high resolution coordinates obtained at room temperature, which allowed us to build an accurate initial model. We have done MD simulations at constant pressure and at constant volume, and we have also maintained a number of atoms with restraints, all of this done to preserve the structure of the crystal.

First, we analyzed constant pressure MD simulations, which can affect the crystal coordinates due to the re-scaling of the volume done by the pressure

coupling mechanism. Second, we fixed the system volume, but the absolute positions of the atoms in this scheme have to be permanently re-adjusted to avoid rotation and translation movement of the whole system. Then, we restrained the position of $44C\alpha$ atoms. This approach solved the previous problem, but generated an over-estimation of the movement of some residues (*i.e.*, the region around residue 75) and an under-estimation of others (the region around the restricted ones). Finally, we restrained the position of only $1C\alpha$, and that is the scheme in which we obtained the better balance of our results, and, in consequence, we recommend it for the study of this kind of protein crystal systems (See Figures 3B and 5).

These strategies allowed us to evaluate the motions of H-FABP in a confined crystalline environment and in solution, observing how the restriction in the atomic positions influences the global motions of the system.

Despite the similarities in the landscape of simulated and experimental B–factors (Figure 5), the edge proteins showed a high fluctuation in some of its residues (see residue Ala75 in Principal Component 1, Supplemental Figure 3 in Supplementary material). However, the unit cells edge volume is well reproduced, indicating that H-FABP packing is described correctly (Figures 3A and 4).

In our analysis, we consider to use the essential dynamics for the calculation of the PCs (Daidone and Amadei, 2012). Since the positional fluctuations are confined to a crystallographic cell, the essential dynamics gives a correct description of the motions when its amplitude is small enough (Supplemental Figures 2 and 3 in Supplementary material). In Table 1, the cross-correlations in the atomic displacements by system indicate collective motion and are, therefore, of potential relevance to H-FABP function (Meinhold and Smith, 2005).

The results presented here are remarkable considering that a direct comparison between X–ray diffraction and MD simulation is difficult, due to the huge differences in the statistical sampling of both techniques.

A typical experimental X–ray data collection is in the order of hundreds of seconds and may involve billions of unit cells. In contrast, in the current computational availability, MD simulations may be extended during microseconds over a small number of unit cells. Despite these limitations in the computational modeling, Molecular Dynamics simulations help us to recover part of the information lost in the experiment, introduce movement and therefore the temporal dimension in the atoms positions, revealing the microstates lost in the averaging process, and let us explore the restrictions to the normal movement of the protein due to confinement. All of this enriches the interpretation of the structure from a biological point of view. The analysis of the normalized correlation coefficients *ncc* reveals a good agreement of the displacement of the protein residues in the simulated conditions, in comparison with the experimental ones, and a better agreement of those simulated in crystalline condition compared to that made in solution.

The main goal of this work was to find a suitable methodology, which can allow us to perform MD simulations of a higher complexity in systems of interest (like the opening/closing mechanism of the H-FABP portal), trusting that chosen scheme captures the essential aspects of the crystalline behavior.

We hope that the methodology to carry out MD simulations of macromolecular crystals that we discuss in detail in this work may contribute to draw attention to the point of protein mobility in time scales shorter than the ones accessible to diffraction techniques, and to clarify it for future studies.

Acknowledgements

Conflict of interest

The authors declare no conflicts of interest.

References

Abraham, M. J. and Gready, J. E. (2011). Optimization of parameters for molecular dynamics simulation using smooth particle-mesh ewald in gromacs 4.5. *Journal of Computational Chemistry*, 32(9):2031–2040.

Abraham, M. J., Murtola, T., Schulz, R., Páll, S., Smith, J. C., Hess, B., and Lindahl, E. (2015). Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, 1:19–25.

Amadei, A., Ceruso, M. A., and Di Nola, A. (1999). On the convergence of the conformational coordinates basis set obtained by the essential dynamics analysis of proteins' molecular dynamics simulations. *Proteins: Structure, Function, and Bioinformatics*, 36(4):419–424.

Amadei, A., Linssen, A. B. M., and Berendsen, H. J. C. (1993). Essential dynamics of proteins. *Proteins: Structure, Function and Genetics*, 17(4):412–425.

Andrec, M., Snyder, D. A., Zhou, Z., Young, J., Montelione, G. T., and Levy, R. M. (2007). A large data set comparison of protein structures determined by crystallography and nmr: statistical test for structural differences and the effect of crystal packing. *Proteins: Structure, Function, and Bioinformatics*, 69(3):449–465.

Batista, P. R., de Souza Costa, M. G., Pascutti, P. G., Bisch, P. M., and de Souza, W. (2011). High temperatures enhance cooperative motions between cbm and catalytic domains of a thermostable cellulase: mechanism insights from essential dynamics. *Physical Chemistry Chemical Physics*, 13(30):13709–13720.

Berendsen, H., Grigera, J., and Straatsma, T. (1987). The missing term in effective pair potentials. *Journal of Physical Chemistry*, 91(24):6269–6271.

Berman, H. M., Kleywegt, G. J., Nakamura, H., and Markley, J. L. (2013). The future of the protein data bank. *Biopolymers*, 99(3):218–222.

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T., Weissig, H., Shindyalov, I. N., and Bourne, P. E. (2000). The protein data bank. *Nucleic Acids Research*, 28(1):235–242.

Bussi, G., Donadio, D., and Parrinello, M. (2007). Canonical sampling through velocity rescaling. *The Journal of Chemical Physics*, 126(1):014101.

Cai, J., Lücke, C., Chen, Z., Qiao, Y., Klimtchuk, E., and Hamilton, J. A. (2012). Solution structure and backbone dynamics of human liver fatty acid binding protein: Fatty acid binding revisited. *Biophysical Journal*, 102(11):2585–2594.

Cerutti, D. S., Freddolino, P. L., Duke Jr, R. E., and Case, D. A. (2010). Simulations of a protein crystal with a high resolution x-ray structure: evaluation of force fields and water models. *The Journal of Physical Chemistry B*, 114(40):12811–12824.

Chen, J., Wang, J., and Zhu, W. (2014). Binding modes of three inhibitors 8ca, f8a and i4a to a-fabp studied based on molecular dynamics simulation. *PLOS ONE*, 9(6):1–10.

Daidone, I. and Amadei, A. (2012). Essential dynamics: foundation and applications. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 2(5):762–770.

Daidone, I., Amadei, A., and Di Nola, A. (2005). Thermodynamic and kinetic characterization of a β-hairpin peptide in solution: An extended phase space sampling by molecular dynamics simulations in explicit water. *Proteins: Structure, Function, and Bioinformatics*, 59(3):510–518.

DeLano, W. L. and Bromberg, S. (2004). Pymol user's guide. *DeLano Scientific LLC, San Carlos, California, USA*.

Friedman, R., Nachliel, E., and Gutman, M. (2006). Fatty acid binding proteins: Same structure but different binding mechanisms? molecular dynamics simulations of intestinal fatty acid binding protein. *Biophysical Journal*, 90(5):1535–1545.

Hess, B., Bekker, H., Berendsen, H. J., and Fraaije, J. G. E. M. (1997). Lincs: A linear constraint solver for molecular simulations. *Journal of Computational Chemistry*, 18(12):1463–1472.

Howard, E. I., Guillot, B., Blakeley, M. P., Haertlein, M., Moulin, M., Mitschler, A., Cousido-Siah, A., Fadel, F., Valsecchi, W. M., Tomizaki, T., Petrova, T., Claudot, J., and Podjarny, A. (2016). High-resolution neutron and x-ray diffraction room-temperature studies of an h-fabp–oleic acid complex: study of the internal water cluster and ligand binding by a transferred multipolar electron-density distribution. *IUCrJ*, 3(2):115–126.

Janowski, P. A., Cerutti, D. S., Holton, J., and Case, D. A. (2013). Peptide crystal simulations reveal hidden dynamics. *Journal of the American Chemical Society*, 135(21):7938–7948.

K. Balendiran, G., Rajendran Pandian, J., Drake, E., Vinayak, A., Verma, M., and Cascio, D. (2014). B-factor analysis and conformational rearrangement of aldose reductase. *Current Proteomics*, 11(3):151–160.

Kondrashov, D. A., Van Wynsberghe, A. W., Bannen, R. M., Cui, Q., and Phillips Jr, G. N. (2007). Protein structural variation in computational models and crystallographic data. *Structure*, 15(2):169–177.

Kruschel, D. and Zagrovic, B. (2009). Conformational averaging in structural biology: issues, challenges and computational solutions. *Molecular Biosystems*, 5(12):1606–1616.

Kuzmanic, A., Pannu, N. S., and Zagrovic, B. (2014). X-ray refinement significantly underestimates the level of microscopic heterogeneity in biomolecular crystals. *Nature Communications*, 5:3220.

Kuzmanic, A. and Zagrovic, B. (2010). Determination of ensemble-average pairwise root mean-square deviation from experimental b-factors. *Biophysical Journal*, 98(5):861–871.

Levin, L. B.-A., Ganoth, A., Amram, S., Nachliel, E., Gutman, M., and Tsfadia, Y. (2010). Insight into the interaction sites between fatty acid binding proteins and their ligands. *Journal of Molecular Modeling*, 16(5):929–938.

Li, Y., Li, X., and Dong, Z. (2014a). Concerted dynamic motions of an fabp4 model and its ligands revealed by microsecond molecular dynamics simulations. *Biochemistry*, 53(40):6409–6417.

Li, Y., Zhang, J. Z., and Mei, Y. (2014b). Molecular dynamics simulation of protein crystal with polarized protein-specific force field. *The Journal of Physical Chemistry B*, 118(43):12326–12335.

Likić, V. A. and Prendergast, F. G. (1999). Structure and dynamics of the fatty acid binding cavity in apo rat intestinal fatty acid binding protein. *Protein Science*, 8(8):1649–1657.

Long, D., Mu, Y., and Yang, D. (2009). Molecular dynamics simulation of ligand dissociation from liver fatty acid binding protein. *PLoS One*, 4(6):e6081.

Ma, P., Xue, Y., Coquelle, N., Haller, J. D., Yuwen, T., Ayala, I., Mikhailovskii, O., Willbold, D., Colletier, J.-P., Skrynnikov, N. R., and Schanda, P. (2015). Observing the overall rocking motion of a protein in a crystal. *Nature Communications*, 6(8361).

Maisuradze, G. G., Liwo, A., and Scheraga, H. A. (2010). Relation between free energy landscapes of proteins and dynamics. *Journal of Chemical Theory and Computation*, 6(2):583–595.

Meinhold, L. and Smith, J. C. (2005). Fluctuations and correlations in crystalline protein dynamics: a simulation analysis of staphylococcal nuclease. *Biophysical journal*, 88(4):2554–2563.

Merritt, E. A. (1999). Comparing anisotropic displacement parameters in protein structures. *Acta Crystallographica Section D*, 55(12):1997–2004.

Olsson, M. H., Søndergaard, C. R., Rostkowski, M., and Jensen, J. H. (2011). Propka3: Consistent treatment of internal and surface residues in empirical pka predictions. *Journal of Chemical Theory and Computation*, 7(2):525–537.

Papaleo, E., Mereghetti, P., Fantucci, P., Grandori, R., and De Gioia, L. (2009). Free-energy landscape, principal component analysis, and structural clustering to identify representative conformations from molecular dynamics simulations: the myoglobin case. *Journal of Molecular Graphics and Modelling*, 27(8):889–899.

Parrinello, M. and Rahman, A. (1981). Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied Physics*, 52(12):7182–7190.

Parthasarathy, S. and Murthy, M. (2000). Protein thermal stability: insights from atomic displacement parameters (B values). *Protein Engineering, Design and Selection*, 13(1):9–13.

Schmid, N., Eichenberger, A. P., Choutko, A., Riniker, S., Winger, M., Mark, A. E., and van Gunsteren, W. F. (2011). Definition and testing of the gromos force-field versions 54a7 and 54b7. *European Biophysics Journal*, 40(7):843.

Stocker, U. and van Gunsteren, W. (2006). Molecular-dynamics simulation of protein crystals: convergence of molecular properties of ubiquitin. In *International Tables for Crystallography Volume F: Crystallography of Biological Macromolecules*, pages 481–488. Springer.

Terada, T. and Kidera, A. (2012). Comparative molecular dynamics simulation study of crystal environment effect on protein structure. *The Journal of Physical Chemistry B*, 116(23):6810–6818.

Tsfadia, Y., Friedman, R., Kadmon, J., Selzer, A., Nachliel, E., and Gutman, M. (2007). Molecular dynamics simulations of palmitate entry into the

hydrophobic pocket of the fatty acid binding protein. *FEBS Letters*, 581(6):1243 – 1247.

Wall, M. E., Van Benschoten, A. H., Sauter, N. K., Adams, P. D., Fraser, J. S., and Terwilliger, T. C. (2014). Conformational dynamics of a crystalline protein from microsecond-scale molecular dynamics simulations and diffuse x-ray scattering. *Proceedings of the National Academy of Sciences*, 111(50):17887–17892.

Xiao, T., Fan, J.-s., Zhou, H., Lin, Q., and Yang, D. (2016). Local unfolding of fatty acid binding protein to allow ligand entry for binding. *Angewandte Chemie International Edition*, 55(24):6869–6872.

Xue, Y. and Skrynnikov, N. R. (2014). Ensemble md simulations restrained via crystallographic data: accurate structure leads to accurate dynamics. *Protein Science*, 23(4):488–507.

Yan, F., Liu, X., Zhang, S., Su, J., Zhang, Q., and Chen, J. (2018a). Electrostatic interaction-mediated conformational changes of adipocyte fatty acid binding protein probed by molecular dynamics simulation. *Journal of Biomolecular Structure and Dynamics*, 0(0):1–13. PMID: 30193557.

Yan, F., Liu, X., Zhang, S., Su, J., Zhang, Q., and Chen, J. (2018b). Molecular dynamics exploration of selectivity of dual inhibitors 5m7, 65x, and 65z toward fatty acid binding proteins 4 and 5. *International Journal of Molecular Sciences*, 19(9).

Yang, L., Song, G., and Jernigan, R. L. (2009). Protein elastic network models and the ranges of cooperativity. *Proceedings of the National Academy of Sciences*, 106(30):12347–12352.

Zamarreño, F., Giorgetti, A., Amundarain, M. J., Viso, J. F., Córsico, B., and Costabel, M. D. (2018). Conserved charged amino acids are key determinants for fatty acid binding proteins (fabps)-membrane interactions. a multi-

methodological computational approach. *Journal of Biomolecular Structure and Dynamics*, 36(4):861–877. PMID: 28298157.

Zamarreño, F., Herrera, F. E., Córsico, B., and Costabel, M. D. (2012). Similar structures but different mechanisms: Prediction of fabps–membrane interaction by electrostatic calculation. *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 1818(7):1691 – 1697.
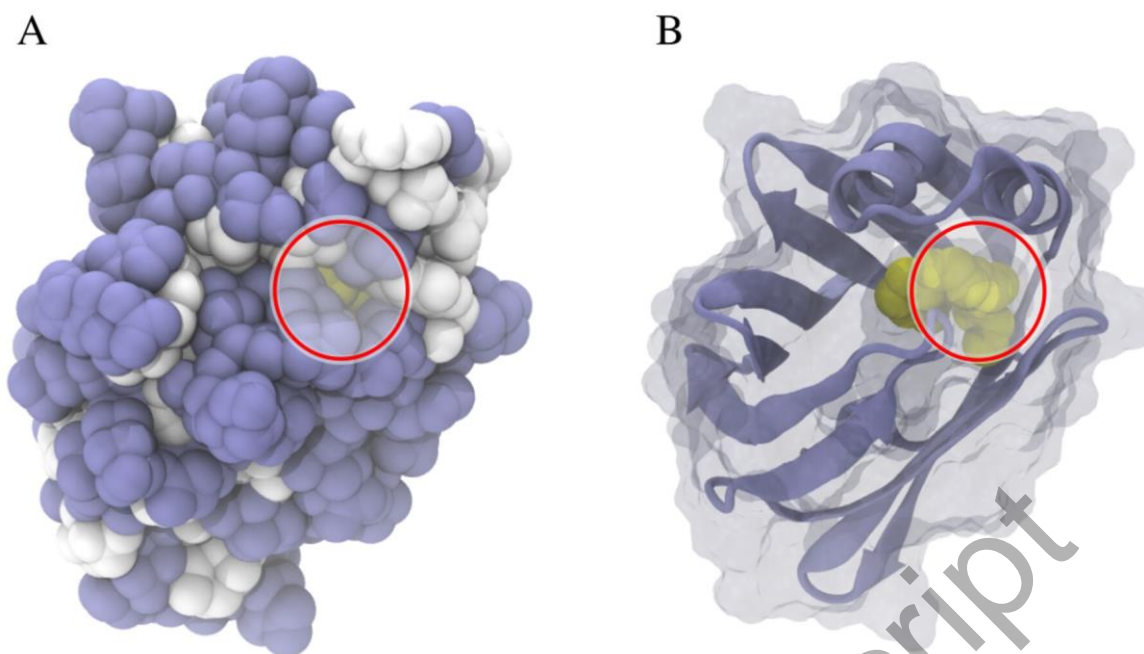
Fig. 1 The ligand can not reach exterior without a conformational change in the protein (PDB ID: 5CE4). In A, protein and ligand are shown in VdW representation, coloured by hydophobic (white) and hydrophilic (iceblue) residues for the protein, and in yellow for the ligand. In B, protein is represented in cartoon (iceblue) and solvent accesible surface (white, transparent), and ligand in VdW (yellow). The only solvent accessible region to the protein interior is highlighted with a red circle, both in A and B.
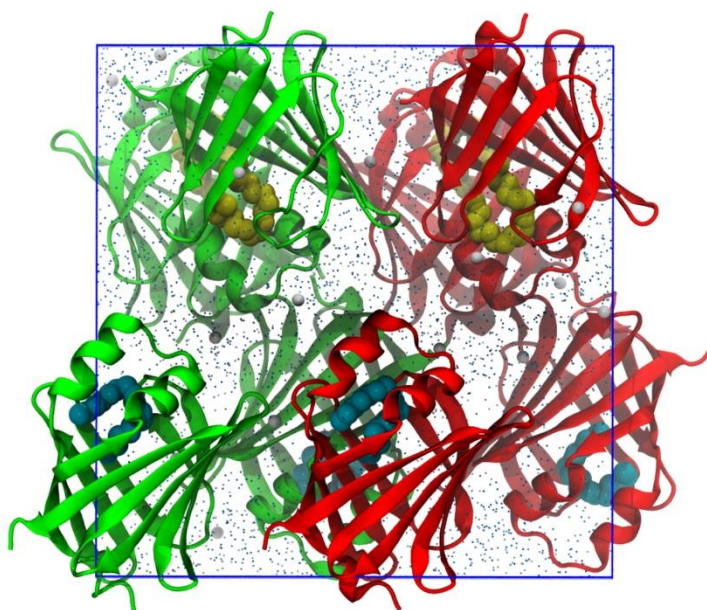
Fig. 2 Initial atomic coordinates for the protein crystal in water. Simulated crystal of the heart fatty acid binding proteins, two unit cells containing four proteins each are arranged in a $2 \times 1 \times 1$ layout. In green and red the four proteins colored by unit cell. The palmitic and oleic acid molecules are represented in cyan and yellow, respectively. Water molecules are indicated using blue spheres. In white spheres Na⁺ ions are represented (for a 3D representation of the system, see the video *crystal.mpg* file and PDB file *crystal.pdb*, incorporated as Supplementary Material).
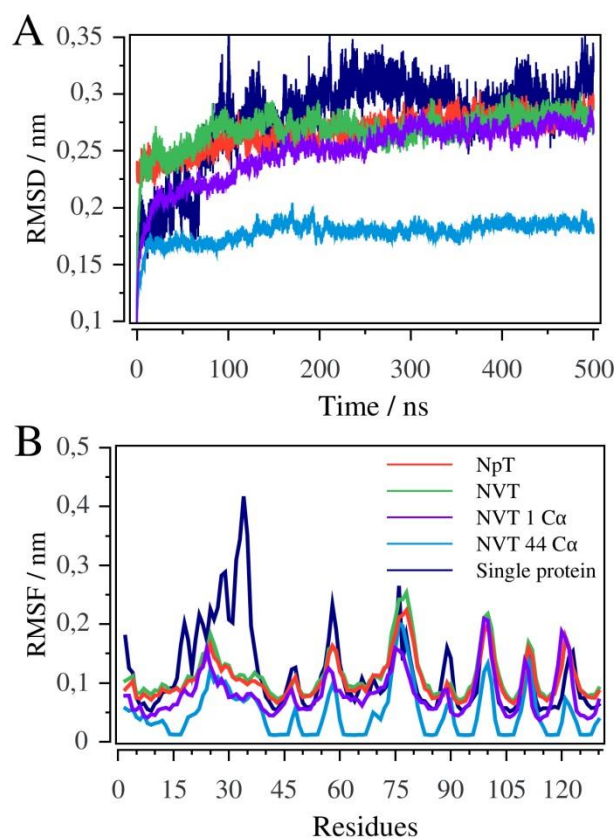
Fig. 3 Structural and mobility analysis of the H-FABP. In A, Root mean square deviation (RMSD) was calculated in a single protein in solution and each of the crystal systems aligning them with the X-Ray and Neutron diffraction structure of H-FABP (Howard et al., 2016). In B, root mean square fluctuation (RMSF) obtained by averaging the eight RMSF curves computed on the $C\alpha$ atoms of each individual crystal proteins and for each $C\alpha$ atoms of the H-FABP in solution. In red, green, violet, light blue and blue lines; crystal NpT, NVT without position restraints, NVT with position restraints ($1C\alpha$), NVT with position restraints ($44C\alpha$), and single protein in solution (NpT), are shown, respectively.
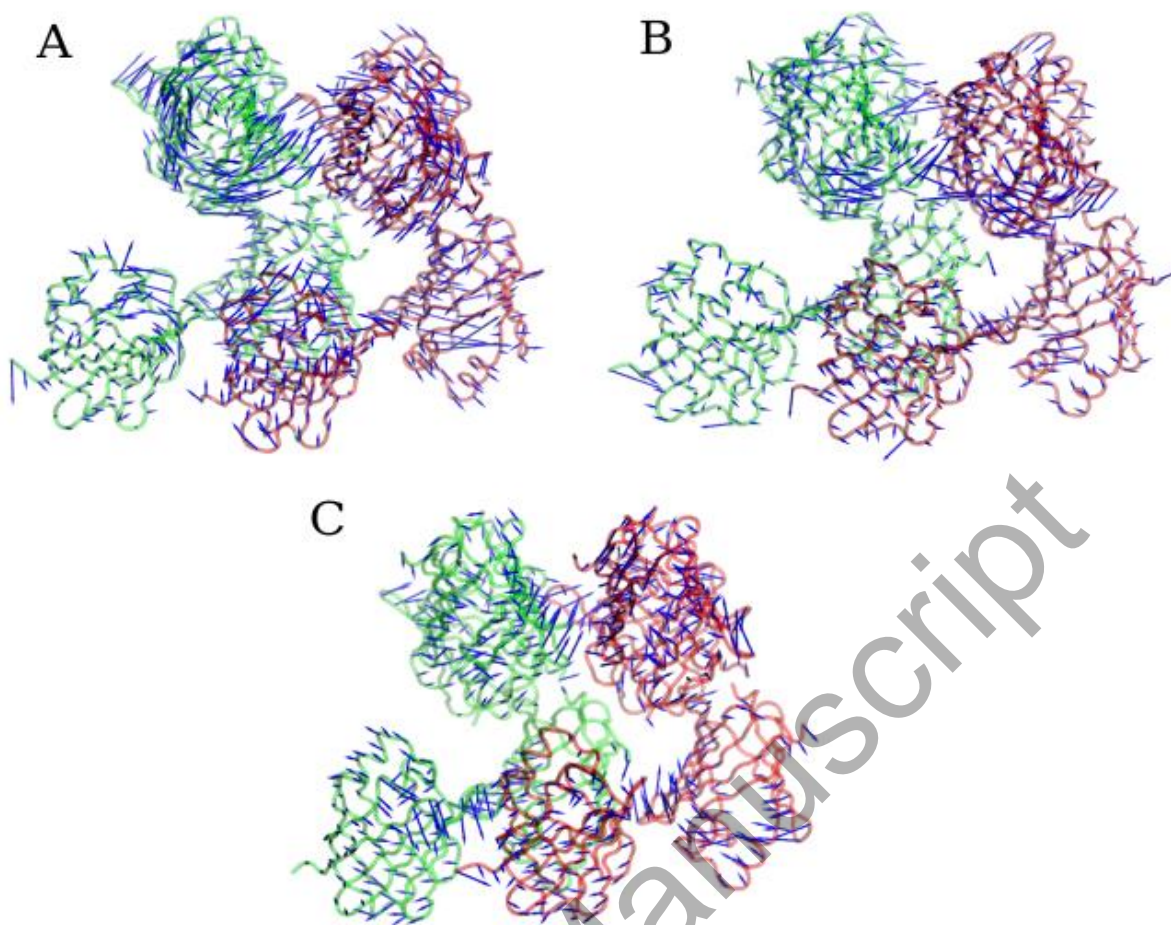
Fig. 4 View of the essential movements in the H-FABP crystal for the first principal component (NVT simulations). The arrows represent the positional fluctuations, their orientation indicating the direction of motion of the Cα atom to which they are attached and their length indicating the amplitude of this motion. In A, B and C, NVT, NVT ($44C\alpha$) and NVT ($1C\alpha$) views are depicted, respectively.
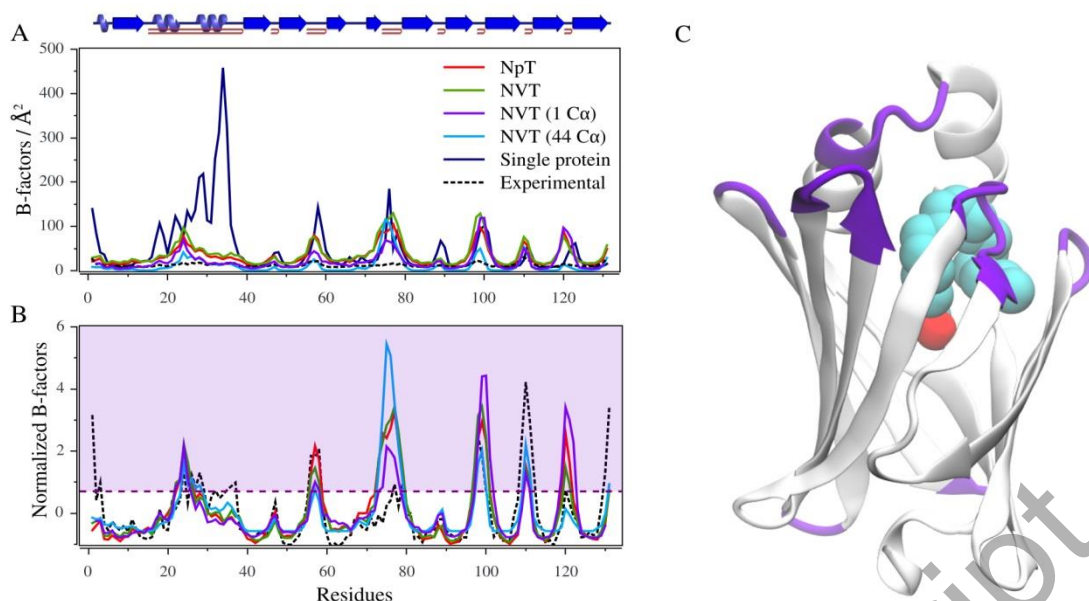
Fig. 5 Simulation B–factors in comparison to experimental B–factors of 5CE4 (Howard et al., 2016). The experimental B–factors are shown as a dash dotted black line. The simulation B–factors were obtained by averaging the eight B–factor curves for the individual proteins in the crystal. Crystal NpT, NVT, NVT with position restraints in 1 Cα and in 44 Cα, and Single protein in solution (NpT), are shown as red, green, violet, light blue, and blue lines, respectively. The Cα atoms were used in this analysis. In A, absolute B–factors are shown for each residue and, in the upper part of the graph, the different regions of the protein with its secondary structure are indicated. In B, normalized B–factors are shown for each residue for all conditions, except for single protein in solution, and the region with normalized B–factors bigger than 0.75 is coloured in violet. In C, a cartoon representation of the secondary structure of the protein initial configuration (5CE4) is depicted, and the residues with normalized B–factors bigger than 0.75 for the NVT (1 Cα) simulation are coloured in violet (*i.e.*, residues 22-25, 57, 74-78, 98-101, 110, 111 and 119-122).
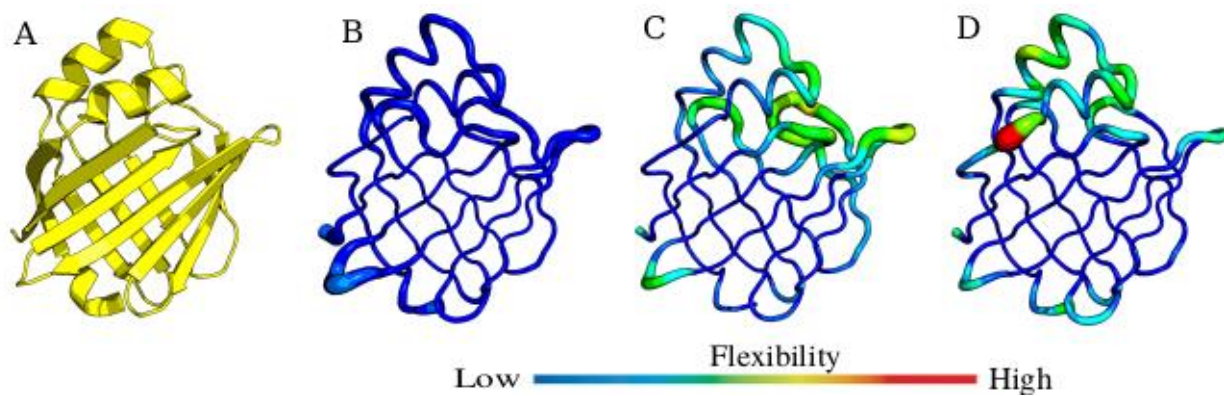
Fig. 6 Isotropic B–factor of H-FABP in NpT ensemble as color code (Flexibility). In A, H-FABP structure from PDB 5CE4. In B, B–factors of experimental H-FABP. C and D (both in NpT ensemble), simulated protein in crystal and simulated protein in solution, respectively. H-FABP structure in A is shown to understand the configurations represented as B–factors.
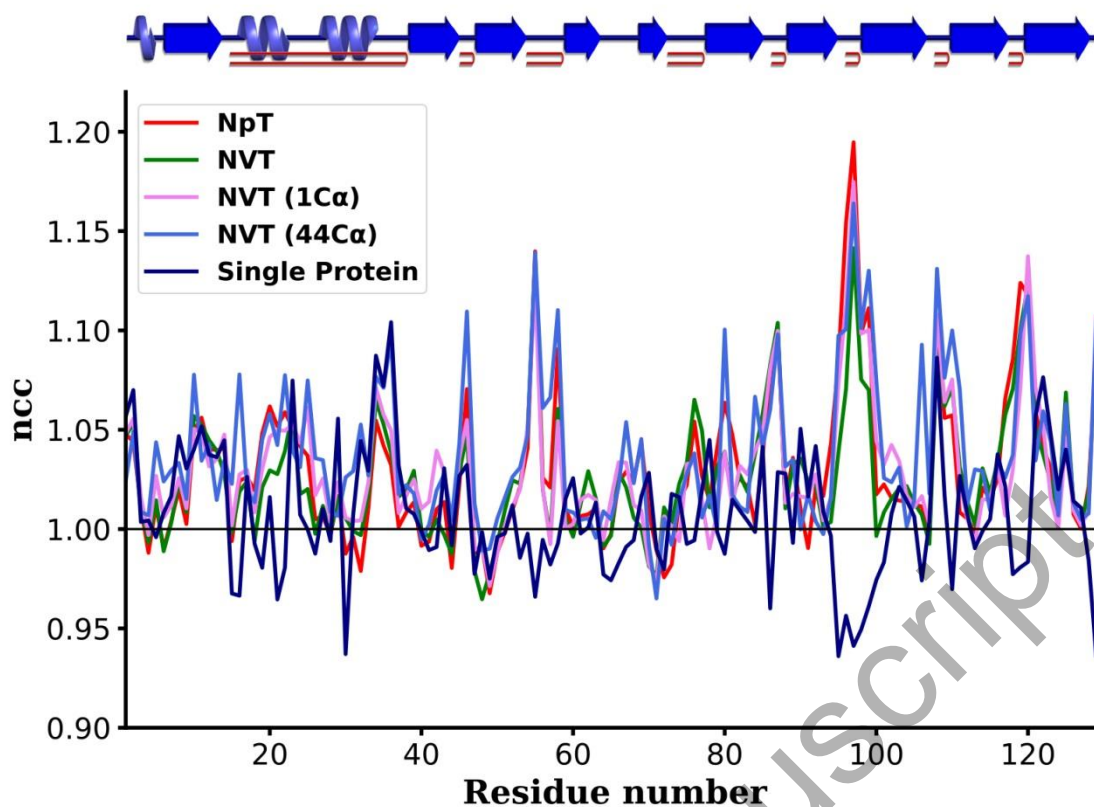
Fig. 7 *ncc* values for all simulated systems, compared to the experimental structure as reference. For the crystal MD simulations, *ncc* values are represented in lines. H-FABP structure is shown over the figure to identify the regions that correspond to each *ncc* value.

Table 1 The root mean square inner products between the one hundred eigenvectors with largest eigenvalues.

Eigenvector system	Root mean square inner product			
	NpT	NVT	NVT($1\,C_a$)	NVT($44\,C_a$)
NpT	1.0	0.761	0.719	0.636
NVT		1.0	0.730	0.630
NVT($1\,C_a$)			1.0	0.627
NVT($44\,C_a$)				1.0

Table 2 Percentage of *ncc* values bigger than 1 for each simulated condition.

	Percentage of *ncc* values > 1
NpT	82.2%
NVT	81.4%
NVT(1 Cα)	90.7%
NVT(44 *Cα*)	93.0%
Single Protein	55.7%