This manuscript will be presented and discussed at a forthcoming Faraday Discussion meeting.
All delegates can contribute to the discussion which will be included in the final volume.

Register now to attend! Full details of all upcoming meetings: http://rsc.li/fd-upcoming-meetings

Faraday Discussions   Volume 164

Tropospheric Aerosol –
Formation, Transformation,
Fate and Impacts

ROYAL SOCIETY
OF CHEMISTRY

ROYAL SOCIETY
OF CHEMISTRY

www.rsc.org/faraday_d

Faraday Discussion 169

ARTICLE

Innovative Interactive Flexible Docking method for multi-scale reconstruction elucidates dystrophin molecular assembly

A.-E. Molza[a], N. Férey[b], M. Czjzek[c], E. Le Rumeur[a], J.-F. Hubert[a], A. Tek[d], B. Laurent[d], M. Baaden[d] and O. Delalande[†,a],

At present, our molecular knowledge of dystrophin, the protein encoded by the *DMD* gene and mutated in myopathy patients, remains limited. To get around the absence of its atomic structure, we have developed an innovative interactive docking method based on the BioSpring software in combination with Small-angle X-ray Scattering (SAXS) data. BioSpring allows interactive handling of biological macromolecules thanks to an augmented Elastic Network Model that combines the spring network with non-bonded terms between atoms or pseudo-atoms. This approach can be used for building molecular assemblies even on a desktop or a laptop computer thanks to code optimizations including parallel computing and GPU programming. By combining atomistic and coarse-grained models, the approach significantly simplifies the set-up of multi-scale scenarios. BioSpring is remarkably efficient for the preparation of numeric simulations or for the design of biomolecular models integrating qualitative experimental data restraints. The combination of this program and SAXS allowed us to propose the first high-resolution models of the filamentous central domain of dystrophin, covering repeats 11 to 17. Low-resolution interactive docking experiments driven by a potential grid enabled us to propose how dystrophin may associate with F-actin and nNOS. This information provides a peak into medically relevant discoveries to come.

Introduction

Dystrophin is an essential protein of skeletal muscle (1) encoded by DMD, the largest human gene. Given the cell membrane ruptures observed when dystrophin is lacking (2), as seen in the lethal genetic disease Duchenne Muscular Dystrophy (DMD), dystrophin function is clearly crucial for muscle survival (3). By interacting with cytoskeleton proteins and membrane lipids, dystrophin enables shear stress resistance in muscle cells (4, 5). Nevertheless, in the absence of a more exhaustive structural characterization, dystrophin's precise role at the molecular level remains poorly understood. The protein's filamentous state and huge size preclude easy answers, but understanding dystrophin at a molecular level, including its structure, dynamics, and interactions, constitutes a key-step in the optimization of gene therapies for rescuing DMD.

Dystrophin is made up of four domains: a N-terminal actin-binding domain; a central rod domain comprising 24 spectrin-like repeats (R1 to R24) (6); a Cys-rich domain involved in the membrane dystrophin-associated glycoprotein complex; and the C-terminal end. These domains are connected by four hinges. The region of repeats 11 to 17 (R11-17) of the central rod domain is involved in crucial interactions with F-actin, nNOS, and membrane lipids (Fig. 1). It has been previously established that the R11-R15 fragment is associated with filamentous actin (F-Actin) (7, 8) and that the R16-17 fragment interacts with neural Nitric Oxide Synthase (nNOS) (9, 10). Both regions interact with membrane lipids (11, 12). However, few structural data are available for the dystrophin protein, and any of these concern the R11-17 region. Interestingly, this region is involved in the hotspot of the in-frame mutations leading to Becker Muscular Dystrophy (BMD) (13). Nevertheless, failure of nNOS recruitment at the membrane is observed in BMD patients, leading to a more severe phenotype (14).

In the present study, we developed a multi-scale reconstruction approach, integrating various experimental and modeling methods. As a proof of concept, the approach was successfully applied, resulting in the structural characterization of the central R11-17 region of dystrophin and its interaction with major protein partners.

With that aim in mind, we first acquired Small Angle X-ray Scattering (SAXS) data from multi-repeat dystrophin fragments. To elucidate the atomic structure of these fragments, we developed theoretical approaches (BioSpring experiments) based on an interactive docking concept (15). This approach uses an augmented Elastic Network Model (aENM) combining the spring network with non-bonded atom or pseudo-atom terms (16). Interactive docking is a user-centred approach that uses advanced 3D interaction devices featuring force feedback (haptics) during a docking simulation to allow application of real-time constraints to the proteins. Details on interactive aspects, haptics, computation of interaction forces and user contributions are provided in Material and Methods and examined in the Discussion section. Complementary to traditional fully-automatic docking approaches, our interactive docking software allows the user to manipulate and drive several flexible proteins to a bound state using her or his own knowledge and expertise of the biomolecular system, but all the while remaining constrained by the underlying biophysical interactions.

We started by acquiring dystrophin structural data using the aENM coupled to the volumetric data provided by SAXS experiments. Thereafter, the macromolecular complex was constructed using the aENM joined to the electrostatic potential grid computed from experimental structures. Interactive driving of the ligand to the receptor was enhanced using contact data obtained experimentally either from footprint-like mapping (trypsin digestion of lipid-protein complexes) or from *in vitro* and *in vivo* cellular experiments on mutant proteins. SAXS-based models provided either ligand (versus F-actin) or receptor (versus PDZ-nNOS) models for dystrophin docking with these two main partners. Beforehand, we performed a rigid low-resolution docking (Attract program) to extensively explore the most probable association modes for the starting structures later considered in the interactive docking step. BioSpring enables both the interactive optimization of the Attract poses following experimental contact mapping, and the evaluation of the effects of introducing flexibility in the molecular models upon their association.

Finally, we present here the first ternary complex composed of F-actin-Dystrophin-nNOS proteins, revealing structural features that could be related to the dysfunction of the truncated dystrophin forms observed in BMD patients.

Experimental

Materials and Methods.

Molecular structures. Similarly to our previous structural study (17), homology models for the R11-15 and R16-17 dystrophin fragments were obtained thanks to the I-TASSER web server (18), using an updated template library that includes the structures of the first repeat of both dystrophin and utrophin (19). The molecular

dynamics (MD) optimized model of the R15-16 fragment was taken from our previous work (17). The actin filaments used to perform the docking experiments were 12-actinG polymers built from the CryoEM structures available in the Protein Data Bank (PDB codes 3MFP and 3G37). The nNOS PDZ crystallographic structure (PDB code 1QAU) was completed using homology modeling to fit Lai's experimental sequence (10). This PDZ structure was submitted to Normal Modes Analysis-driven deformations using the elNémo server (20, 21) to produce several conformations. All structural data are shown in a schematic cellular context in Figure 1.

Low-resolution rigid docking. Rigid low-resolution Attract docking calculations were performed for both R11-15—F-actin and PDZ-nNOS—R16-17. The most relevant associations between the structures, that could be further associated interactively, were predicted through Attract exhaustive searches. Newly-developed parameters for the Attract force field were used during the docking calculations (22). Placing of the ligands, R11-15 or PDZ-nNOS for each experiment, was achieved thanks to density grids of 25Å and 10Å. A total of 157 starting points for the dystrophin R11-15 fragment were set all around the actin filament, creating a 140Å-high cylinder with a 150Å radius (see Supporting Information). The PDZ-nNOS had 381 starting points around R16-17. For each starting point, the standard 258 rotations were computed for the ligand. Final data was a result of 40,506 poses and 98,298 poses respectively for the R11-15—F-actin and PDZ-nNOS—R16-17 docking experiments. Ranking of the best poses was performed by sorting by complex energy and clustering over the ligand position with a root mean square deviation (RMSD) of 60Å (F-actin docking) and 10Å (PDZ-nNOS). Best theoretical complexes were thus selected for their low energy levels (Attract function) and because they were the most populated clusters. Molecular contacts were defined with a cutoff distance between ligand and receptor of less than 5Å.

Augmented Elastic Network. The coarse graining of initial homology models and experimental PDB structures was performed in accordance to the rules defined by Zacharias (23). The augmented Elastic Network Model (aENM) combines the usual ENM model (harmonic functions) and non-bonded terms (Attract coarse-grain force field) as per our previous work (16) and similar to that proposed by Periole et al. (24) or more recently by Globisch et al. (25). The aENM model supports the use of multiple cut-off distances to designate a final spring network (Fig. 2A). It thus allows the assignation of variable weight to secondary or ternary structures as needed. In this study, a cut-off distance of 9Å was used for backbone particles involved in secondary structure elements, and it was reduced to 5Å for other particles that belong to loops or side chains. The two resulting spring networks were merged into a final aENM model after the particles of both previous aENM models were linked by a spring if they were closer than the cut-off distance of 5Å. All-atom aENM was used for high-resolution reconstruction of the coarse-grain models. Double cut-off ENM had distances of 9Å for backbone atoms and 3Å for side chain and loop atoms, and 3Å was used as the distance cut-off to merge the first two ENM models. Atomic parameters, including charge and radius, were extracted from the AMBER99 force field (26).

Potential grid. Electrostatic potential maps were obtained from APBS calculations (27) achieved for the structure of the receptor molecules (actin or dystrophin R16-17 in each docking experiment) using the Amber force field and 50 mM sodium and chloride ions. APBS-provided electrostatic potential grid data in the OpenDX format are compatible with the BioSpring reading routine.

Experimental volume maps. SAXS scattering curves were analyzed using the ATSAS software suite (28). Twenty molecular shapes were generated using Gasbor's method which is based on Monte-Carlo draws of hard-sphere distributions (one bead per amino acid) paired with cross-validation via scattering curve simulation. The Oligomer program was used for weighing the 20 shapes (linear combination solution). Oligomer ranking showed that the best SAXS molecular shapes selected for both R11-15 and R16-17 fragments achieved a high rank (#1 for both) and high weight percentage (43% and 31% contribution to the scattering curve, respectively) as well as a low $\square^2$ value (1.3 and 2.5, respectively). The volumetric data for the Gasbor molecular shapes were converted into a pseudo-density grid using the MDFF applet of the VMD program (29) at a resolution of 20Å.

Interactive human-guided simulations. As in our previous studies (16, 30), the MDDriver library was used for data exchange between the simulation engine, in this case BioSpring calculations, and the visualization frontend, here the VMD viewer. The interactive simulations were computed on a 24-core desktop computer using a PHANTOM Omni haptic device (SensAble) for advanced 3D interaction. The 3D aspects of this device make it easier to pick an atom or a coarse-grained particle within the protein, which is intrinsically a spatially complex 3D object. Haptic features allow the user to feel the forces involved in the simulation through force feedback (see next paragraph for a more detailed discussion). The BioSpring software manages the aENM model and guides it through a scalar density field (read from data in the OpenDX format) by computing gradients to derive the force to be applied to a particle according to its mass (SAXS grid) or charge (APBS grid). As these simulations are user-guided, they need to be repeated several times in order to account for the variability of the interaction.

Haptic interactions. To provide haptic feedback to guide the user, forces must be computed and rendered at very high frequencies, usually between 300Hz and 1000Hz. This framerate is clearly incompatible with typical molecular simulation framerates, especially for complex structures, because each timestep is computed at a relatively low frequency, at most still an order of magnitude below the haptic requirements. To overcome this issue, an intermediate high frequency "simulation" is asynchronously used to provide haptic feedback. It consists in simulating springs virtually connecting the avatar of the haptic probe device in the simulation space and all the particles selected by the user. A sum of the spring forces is then rendered to the user through the haptic device, and each force on the particle is sent to the actual BioSpring simulation engine at each timestep as an external simulation constraint. The user-felt forces thus indirectly originate from the molecular simulation, via the connecting springs. This solution provides usable and useful haptic feedback. As the perception system of the user naturally integrates force feedback over time, (s)he is able to feel the flexibility or rigidity of the area that is haptically manipulated. Likewise, using springs allows the user to finely tune the global stiffness of these springs to render the bonded and non-bonded interactions computed in the molecular simulation more accurately.

Interactive docking experiments. A strict protocol was set up and used for the six experiments carried out for each scenario (envelope, orientation) to prevent *a priori* bias that might occur for the interactive fit simulations of driving models into SAXS volumetric data (for instance placing a specific coiled-coil repeat in a chosen spot of the SAXS envelope). Predefined backbone particle selections of each alpha-helix were used to translate the aENM model into the volumetric grid. The first step of the protocol consisted in the orientation of the aENM filament parallel to the molecular envelope.

The entire aENM was then driven into the envelope. The aENM model was then stretched beyond both terminal ends of the molecular envelope, then was let to relax on is own, without interactive intervention, into the volumetric grid and over 20,000 calculation steps. In the case of interactive docking driven by an APBS potential grid, starting positions for the ligand were either set randomly or by using one of the best poses predicted from the preliminary Attract docking calculations. To complete efficient interactive simulations at an all-atom resolution on desktop computers, the OpenCL implementation newly developed by us for the BioSpring software was used.

Structure validation and representation. Validation of SAXS-based coarse-grained models was performed by volume overlap calculations in the Chimera program (31). The final atomic models were evaluated via the SAVeS metaserver (http://services.mbi.ucla.edu/SAVES/), which combines different programs for molecular structure quality control. All figures were realized with VMD or Chimera.

Results

Interactive fit of homology models into SAXS volumetric data: a quick and efficient method for producing reliable SAXS-based models.

An interactive volume fitting procedure (Fig. 2B) enabled us to reconstruct the dystrophin central domain in less than a minute. For the short fragment, a SAXS-based model was produced without human-added driving force pushing the aENM into the volumetric data. In this particular case, haptic rendering alone provided essential information about the rapid convergence of the fit during the simulation. The good quality (see Supporting Information) of the models built following the interactive fitting procedure is based on two main features. Firstly, the robustness of the aENM allows us to maintain the secondary structure as well as the particle packing, thus avoiding the collapse of the structure so that reliable solutions could finally be obtained. Indeed, the basic fold of spectrin-like repeats in triple coiled-coil (17, 19) was maintained for each dystrophin repeat model. The flexibility of the model still enables individual alpha helices to kink or coiled-coil repeats to rotate towards each other as needed. Secondly, online visualization allows a good evaluation of the volumetric fit between the aENM and the SAXS envelope. This enables the user to discard bad subdomain placement in the experimental volumetric data. It notably made it possible to discriminate between the different N-terminal to C-terminal or C-terminal to N-terminal orientations stemming from the absence of linear shape polarisation of the SAXS volumetric data. The final interactive reconstruction of a SAXS-based protein structure model was done by superimposing the all-atom homology model (aENM format) onto the final aENM position obtained by the preliminary coarse-grain interactive fitting, then by calculating the flexible volumetric fit with the all-atom aENM. The final volume overlaps obtained for the R11-15 dystrophin fragment were respectively 60%, 86%, and 92% for the static homology model, flexible coarse-grain model, and flexible all-atom model. Overlap percentages for the R16-17 fragment were 84% (homology), 87% (coarse-grain aENM), and 91% (all-atom aENM).

Given the fact that the R15-16 dystrophin fragment is unable to be produced *in vitro*, thus precluding SAXS data acquisition, we used the previously obtained MD models (17) to merge the two SAXS-based all-atom models and to propose the first model of the R11-17 fragment of dystrophin. Final SAXS-based atomic models revealed the doglegged form of dystrophin's filamentous structure. It is worth noting that alpha-helical structures and coiled-coil arrangements can be maintained despite the sharp kinks between several consecutive repeats (for instance between R14 and R15) that have been observed experimentally (SAXS molecular shapes).

Thanks to the interactive exploration and evaluation of the two possible relative orientations (N- to C-terminal end or C- to N-terminal end) within the SAXS envelopes, we found that the largest kinks are located at the R12-13, R14-15 and R16-17 junctions (linker regions). These linkers are either characterized as straight ☐lpha-helices or as helices bearing a kink of its main axis. Indeed, SAXS volumetric data correspond to a time-averaged molecular shape, and the kinks we observed should correspond to a mean position between two existing extreme conformations. In fact, kinks could have the wrong amplitude but impose the correct relative direction between two consecutive repeats. These flexibility points were considered as "preferential adjustable regions" for the further dystrophin-F-actin docking interactions.

Interactive fit of SAXS-based models into APBS potential grids: how to create a theoretical complex using guidelines produced by systematic low-resolution docking combined with footprint-like experimental data.

We explored the association modes between dystrophin and F-actin by analyzing the two available isoforms of filamentous actin when associated with ADP (PDB-id 3MFP) or with ADP, PO_4^{2-} and Mg^{2+} (PDB-id 3G37). We considered these two conformational states to be representative of changes that could occur *in vivo* in a stretched or relaxed muscle cell, and each was used as static receptor conformation for the interactive docking experiments.

The automated preparatory low-resolution Attract docking experiments provided evidence for the most probable F-actin molecular surfaces that could interact with the R11-15 SAXS-based model. A parallel orientation of both filaments seemed favoured, and the dystrophin fragment is inserted in a F-actin hydrophobic groove and surrounded by small electronegative patches. This latter point is in agreement with the observation that the R11 to R15 repeats are among the most positively-charged dystrophin repeats (17). It appears that the three best complexes obtained by Attract, independently from the F-actin structures, i.e. with or without Mg^{2+}, showed very similar contact patterns. The first type of association (clusters ranked #1 and #2) showed preferential contacts between dystrophin's R14 and R15 repeats and F-actin, while R11 to R13 showed only poor contacts with the actin filament. The second association mode (clusters ranked #3) highlights preferential contacts between R11 and R12 repeats and F-actin. We completed contact frequency mapping by recording the contacts between both proteins with an inter-particle distance ≤ 5 Å in the first 20 molecular complex clusters obtained by the rigid docking search. We used this mapping to visualize the most probable contact surface on the F-actin. This first docking using Attract did not allow dystrophin flexibility and likely avoids a

proper contact between the two filaments, which would explain the discrepancy between the two best solutions with two different regions of dystrophin.

Next we therefore used a flexible and human-guided docking approach with BioSpring. The driving force in interactive docking with BioSpring was provided by an APBS electrostatic potential grid computed on the F-actin experimental isoforms. This docking strategy seems highly relevant because of the predicted importance of electrostatics in the interaction of dystrophin with F-actin (12). As an APBS map is calculated for experimental static structures, we considered the F-actin to be a static molecular shape, thus avoiding the miscorrelation in the spatial distribution of volume and electrostatics that would occur if we had proposed a flexible model (aENM) for the receptor filament. Interactive BioSpring docking enabled a flexible adaptation of dystrophin to the static F-actin molecular surface following the identified interaction groove predicted by the Attract-provided static low-resolution search. As shown in Figure 3, graphical rendering of the molecular surface potentials of F-actin during the interactive docking proved to be a powerful help in the interpretation of the dynamic behaviour of the R11-15 dystrophin ligand driven into the potential grid of the F-actin receptor.

The interactive flexible simulation enabled us to obtain a macromolecular complex in agreement with experimental data. In fact, we had previously (33) performed a MSMS analysis after trypsin digestion of R11-15 in interaction with liposomes: the peptide fragments accessible to trypsin proteolysis were predicted to be free of phospholipid interactions and therefore to be potentially involved in another interaction, such as with F-actin. The conclusion then was that helices B and C of repeat R13 and helices A and B of repeat R14 could be involved in an interaction with the actin filament.

We selected the two best association modes proposed by Attract docking as our starting point for the interactive docking simulations. Using these, it appeared that matching the experimental contact restraints would impose a change in the R12 and R13 positioning towards the F-actin filament (Fig. 3). Indeed, in all the Attract poses, a contact between helix C of R12 is observed and should be avoided, as the region is expected to preferentially interact with sarcolemma phospholipids. The torsion of the R12-13 linker region allowed us to interactively match the previously-observed experimental contacts, and led to a decrease in the kink angle observed in the SAXS-based model. Looking at the macromolecular complex, it appears that dystrophin showed a modified topography when compared to the SAXS-based model obtained in solution, as the succession of repeats from R11 to R13 show a rather linear filamentous structure. On the contrary, the kink observed at the R14-15 linker region in the SAXS-based model of the R11-15 fragment was maintained in the dystrophin–F-actin complex. Interactive docking led to a continuous contact of both filaments on a dystrophin track defined from R11 helix B to R12 helix A, and from R13 helix B to R14 helix C. These results are in agreement with the experimental data collected earlier.

Interactive fit of SAXS-based models into APBS potential grids: how to create a theoretical complex using guidelines obtained from systematic low-resolution docking and combined with mutagenesis experiments.

The low-resolution and rigid explorations of the association between nNOS structures and the R16-17 SAXS-based models showed three preferential relative orientations of the PDZ towards the dystrophin fragment (ranked #1, #2 and #6). All three displayed a good correlation with the contact surface identified indirectly on dystrophin by our recent theoretical study (34) and through the mutagenesis experiments performed *in vitro* and *in vivo* by Lai and coworkers (10). In comparison to this contact data, the fraction of native contacts in the molecular complexes obtained here are respectively 53%, 43%, and 60%. However, even if the PDZ surface in contact with the R16-17 fragment seemed well-defined, the relative orientation of the nNOS PDZ subdomain towards the dystrophin fragment itself remained unfocused, with two of these poses head-to-tail.

The interactive docking experiments associating the nNOS PDZ subdomain to the R16-17 dystrophin fragment (Fig. 4) used the protocol previously detailed for dystrophin/F-actin docking, but with a major change. This time, in addition to dystrophin's flexibility, we also considered the relative flexibility of the receptor (R16-17), as both proteins are represented by an aENM. To avoid sliding or rotation of the receptor aENM model towards its APBS potential grid, the terminal-end C☐ atoms of the dystrophin fragment were fixed (spatial restraint). The use of such a rigid-flexible heterogeneous aENM model enabled, upon interactive docking, a smooth modulation of the kink characterized experimentally in the SAXS shape at the R16-17 linker region. Long-range driving of the PDZ ligand to the target surface highlighted experimentally is not trivial given the poor contribution of electrostatics expected in the association of the two proteins (34). In these conditions, the three best poses produced by rigid docking were used as starting positions for the PDZ subdomain. Therefore, interactive experiments operate as a precise exploration of the alternative poses, allowing for rapid and accurate convergence on the most probable association mode, completed with online visual checks. The flexible adaptation of both molecular surfaces to each other was observed in each interactive simulation. This interactive flexible approach enabled a better match of the complex initially ranked #6 by the Attract function score. The fraction of native contacts in this final molecular complex is 80%, corresponding to an increase of 20% over the corresponding Attract pose. This result underlines the usefulness of flexible modeling in the docking approach.

Upon the interactive exploration of the three best alternative Attract poses, we tried to rank them by perturbation/relaxation trials. Online visualization of the docking simulation confirmed the PDZ pose ranked #1 as being poorly stabilized despite its good Attract score. The pose ranked #6 by Attract was selected by BioSpring as the most probable lowest energy position of the nNOS subdomain towards the R16-17 dystrophin fragment. Pose #2 was observed as an intermediate stage between pose #1 and #6. Indeed, it could be highly relevant to hypothesize that the two less stabilized poses are transient orientations of the PDZ enabling it to reach its best association mode with dystrophin R16-17.

Interactive fit of high-resolution structures into pseudo-density data: a generic and efficient method to perform all-atom reconstruction of coarse-grain models.
Final all-atom reconstruction of the full ternary complex was completed either (i) by a fragment-building strategy using a lowest-RMSD superimposition onto the coarse-grained models previously obtained, followed by a short energy minimization step, or by (ii) a volumetric data fit using an all-atom aENM positioned into a grid derived from the coarse-grained models (Fig. 5A). The second approach turned out to be the most efficient for yielding reliable models in the shortest time. This quick coarse-grain to all-atom reconstruction method using BioSpring enabled us to propose structurally-stable high-resolution models of dystrophin-F-actin and nNOS-dystrophin complexes obtained from low-resolution interactive docking simulations. The ternary macromolecular complex was obtained by superimposition (lowest RMSD on C☐) of these two atomic models onto the SAXS-based R11-17 dystrophin model (Fig. 5B).

Discussion

Exploring molecular plasticity: modeling biomolecules under the control of experimental data as if they were clay.
Using the BioSpring software as a powerful model builder enabled us to confirm the robustness of the aENM approach to a user-imposed local deformation of the driving and shaping of biomolecules through interactive simulation. Molecular object inertia or heterogeneous response to the forces applied uniformly to a set of particles can be overcome by pre-selecting the interactive ensemble. An approach consisting of preferential interaction on a selection of protein backbone particles leads to less deformed structures and to a better haptic response, as the particles are more closely connected due to the higher particle-density in their neighbourhood. The models' quality was further enhanced by a multi-cut-off spring-network design avoiding molecular distortion to a very high degree. We verified that the non-bonding terms (steric and electrostatics) set in the aENM model efficiently counterbalance the tendency of ENMs to remain very close to their initial conformation. For instance, the flexibility of the R11-15 model enabled us to induce conformational changes to dystrophin upon its complexation with F-actin.

We should emphasize that molecular packing and folding is not only rendered by the steric term associated to a particle's volume, but structurally encoded by springs in the aENM. Steric clashes, and more seriously structural collapse, are thus easily avoided upon molecular docking or volumetric data fitting when compared to what can be observed with other fitting methods such as MDFF (35). BioSpring interactive docking into volumetric data enables the calculation of either intra-molecular or inter-molecular interactions using non-bonded terms of the aENM representation, in contrast to a simple volume match guided by a density map. This representation considerably improves the quality of the flexible fit under experimental volumetric restraints, notably aiding in the judicious placing of loops in biomolecular models. We tested BioSpring experiments as an efficient and quick generic method for multi-scale reconstruction by generalizing the approach, combining flexible molecular models and a spatial distribution fitting frame. The speed reliability and convergence of fitting the models into the volumetric target also make an offline procedure possible. This is comparable to the recently published Internal Coordinates Flexible Fitting (ICFF) method (36), which is several orders of magnitude faster than a Molecular Dynamics approach such as MDFF. However, detailed information about flexible regions is required before using the ICFF method to its full efficiency. Our more generic

aENM-based approach allows limited flexibility in amplitude but does not necessarily impose the set of a pre-defined flexible ensemble, since the regions bearing an increased flexibility are intrinsically related to regions with lower density of particles. Furthermore, BioSpring is quicker than MDFF of a corresponding atomistic model thanks to coarse-graining and code optimization (OpenCL).

Long-range driving of the ligand towards the receptor was essential to assist the scientist's gesture through haptic rendering. A possible limitation in electrostatic grid guidance during the interactive docking experiments could arise either from biomolecular system size and complexity or from a lack of importance of electrostatics in the association of the two studied partners. Displaying experimental contacts on the ligand and/or the receptor provides essential cues for the interactive docking procedure, revealed to be more rapid and more accurate than without them. Exploration of space for the approach of the ligand towards the receptor thus could be reduced until less than twenty per cent of an exhaustive positioning by an automated method. By using an exhaustive low-resolution docking method prior to the interactive procedure, we prevented the oversight of putative association modes. The combination of exhaustive-computer and knowledge-based human approaches led to reach efficiently the best solution in accordance to the interaction surfaces defined by mutation or footprint experiments.

The interactive procedure enabled us also to drive the association by proposing complex moves that are difficult or impossible to set up with traditional or offline docking approaches. For instance, during dystrophin-actin docking, the rotation of a part of the filament around its main axis (R12-13) was necessary to achieve a reliable match with data proposed by experimental contact mapping. Moreover, distancing a ligand subdomain from an unwanted position in the proximity of the receptor, as performed for the R15 repeat towards F-actin, is not easy to achieve in traditional docking programs, usually because they intrinsically force the matching of both the ligand and the receptor.

It is logical to note that the correlation between the experimental data increased if the flexibility was also extended from the ligand to the receptor contact surface, simply keeping its backbone fixed. The modeling of biomolecular matter can be easily set up by the design of a rigid/flexible heterogeneous construction from the BioSpring software commands, which distinguish between static and dynamic particle sets. This BioSpring option enabled us to considerably speed up the interactive calculation rate for very large systems, with the simulation being faster in proportion to a smaller size of the dynamic particle set.

Overall this mixture of human interaction based on pre-calculated guidelines with experimental data and an underlying physical model has proven very efficient. It remains however challenging to characterize, describe and quantify the human contribution that may vary from one run to another. Therefore several runs are systematically carried out, until a convergence or a consensus is achieved. Haptics make the interaction natural, as would be the manipulation of a physical model, e.g. made of clay. In particular guidance by the interaction forces, firstly electrostatics, is an invaluable cue. Human (and pre-calculated) knowledge allows the user to significantly reduce the search space and is critical to supervise the refinement of the models. This raises the question whether the solutions that were found are unique. The subjective assessment of the authors is that no other – significantly different – models compatible with the experimental constraints could be built with the underlying physical model.

Exploring the dystrophin's interactome: modeling the protein complex formed in the region of a hotspot of a genetic pathology. In the final ternary complex, three major observations brought crucial understanding to the macromolecular system involving dystrophin. The first concerns the weak contacts of the R15 repeat with F-actin and of the R16 repeat with nNOS. Those two repeats seem to be important for structuring the dystrophin filament, but also emphasize a structural frontier between the contact positions of each protein partner. This may be related to the atypical feature of the R15-16 linker region, a loop twelve-residues long which is probably at the origin of its *in vitro* instability. The second observation highlighted by our macromolecular model deals with the relative positioning of dystrophin and actin filaments with respect to each other. While the N-terminal end of dystrophin R11-15 fragment is laid out parallel to the F-actin, mainly interacting with only one of the two actin filaments, the C-terminal end kinks at the linker regions of the final repeats, disengaging from F-actin. This contact pattern extends along a distance of 150 Å and with a coiling of both dystrophin and actin filaments over an angle of 180°. This observation is compatible with both of the F-actin isoforms considered in our study, and indicates that dystrophin does not seem to bridge the two dimerizing filaments of F-actin. The third main observation concerns the relative positioning of F-actin towards the nNOS PDZ subdomain. In the final macromolecular complex, the dystrophin R11-17 contact surfaces interacting with F-actin and nNOS are oriented in a similar manner. This could emphasize the structuring of another interaction subdomain, that of the potential regions in contact with phospholipids. In the truncated BMD forms of dystrophin, this polarised 3D organisation could be affected, resulting in the loss of efficiency in dystrophin's essential role as a scaffold protein for the muscle cell membrane.

Conclusions

The interactive flexible method based on the BioSpring software presented here allows one to easily swap from a one-residue coarse-grain level (SAXS molecular shape in the Gasbor method) to the Attract coarse-grain level, finally coming up with a high-resolution molecular description. The approach is a highly reliable multi-resolution model builder for the construction of large molecular assemblies. In the future, the interactive simulations of the fitting of molecular models to SAXS data could be greatly improved by an instant display of the scattering curve calculated for a given state of the aENM.

We present here the first models of dystrophin fragments in solution, based on SAXS measurements. We also associate, using an interactive flexible docking approach, the dystrophin R11-15 SAXS model with the F-actin atomic structure, obtaining a final macromolecular complex that correlates well with experimental contact mapping. We also propose the first molecular model for the association of dystrophin R16-17 to the nNOS PDZ subdomain, which is well supported by *in vitro* and *in vivo* data. The ternary complex involving the BMD hotspot region of dystrophin (R11-17) with two of its main protein partners was modeled by combining theoretical and

experimental information, leading to a highly-reliable structural model.

Acknowledgements

Funding

Availability

The BioSpring software is an open source program available online at http://sourceforge.net/projects/biospring.

Notes and references

[a] Institut de Génétique et Développement de Rennes, CNRS, UMR6290, Univ. Rennes 1, Campus Santé, 2 av du Pr Léon Bernard, 35043 Rennes Cedex, France.

[b] Laboratoire d'Informatique pour la Mécanique et les Sciences de l'Ingénieur, CNRS, UPR3251, Univ Paris XI, Bât 508, 512 et 502 bis, 91403 Orsay Cedex, France.

[c] Laboratoire Végétaux Marins et Biomolécules, CNRS, UMR7139, Univ. Pierre et Marie Curie-Paris 6, Station Biologique, 29682 Roscoff, France.

[d] Laboratoire de Biochimie Théorique, CNRS, UPR9080, Univ. Paris Diderot, Sorbonne Paris Cité, 13 rue Pierre et Marie Curie, 75005 Paris, France.

† corresponding author.

Electronic Supplementary Information (ESI) available: Figure representing Attract docking starting points and table for the SAVES structural analysis are accessible as supplementary data. See DOI: 10.1039/b000000x/

1 M. Koenig, A. P. Monaco and L. M. Kunkel, Cell, 1988, **53**, 219.

2 B. J. Petrof, J. B. Shrager, H. H. Stedman, A. M. Kelly and H. L. Sweeney, Proc Natl Acad Sci U S A, 1993, **90**, 3710.

3 M. Koenig, E. P. Hoffman, C. J. Bertelson, A. P. Monaco, C. Feener and L. M. Kunkel, Cell, 1987, **50**, 509.

4 E. Le Rumeur, J. F. Hubert and S. J. Winder, FEBS Letters, 2012, **586**, 2717.

5 E. Le Rumeur, S. J. Winder and J. F. Hubert, Biochim Biophys Acta, 2010, **1804**, 1713.

6 S. J. Winder, T. J. Gibson and J. Kendrick-Jones, FEBS Letters, 1995, **369**, 27.

7 K. J. Amann, B. A. Renley and J. M. Ervasti, J Biol Chem, 1998, **273**, 28419.

8 I. N. Rybakova, K. J. Amann and J. M. Ervasti, J Cell Biol, 1996, **135**, 661.

9 Y. Lai, G. D. Thomas, Y. Yue, H. T. Yang, D. Li, C. Long, L. Judge, B. Bostick, J. S. Chamberlain, R. L. Terjung and D. Duan, J Clin Invest, 2009, **119**, 624.

10 Y. Lai, J. Zhao, Y. Yue and D. Duan, Proc Natl Acad Sci U S A, 2012, **110**, 525.

11 S. Legardinier, C. Raguenes-Nicol, C. Tascon, C. Rocher, S. Hardy, J. F. Hubert and E. Le Rumeur, J Mol Biol, 2009, **389**, 546.

12 J. Sarkis, V. Vié, S. J. Winder, A. Renault, E. Le Rumeur and J. F. Hubert, FASEB J, 2013, **27**, 359.

13 S. Tuffery-Giraud, C. Béroud, F. Leturcq, R. Ben Yaou, D. Hamroun, L. Michel-Calemard, M. P. Moizard, R. Bernard, M. Cossée, P. Boisseau, M. Blayau, I. Creveaux, A. Guiochon-Mantel, B. de Martinville, C. Philippe, N. Monnier, E. Bieth, P. Khau Van Kien, F. O. Desmet, V. Humbertclaude, J. C. Kaplan, J. Chelly and M. Claustres, Hum Mutat, 2009, **30**, 934.

14 C. Gentil, F. Leturcq, R. Ben Yaou, J. C. Kaplan, P. Laforet, I. Penisson-Besnier, C. Espil-Taris, T. Voit, L. Garcia and F. Pietri-Rouxel, Hum Mol Genet, 2012, **21**, 3449.

15 A. Tek, M. Chavent, M. Baaden, O. Delalande, P. Bourdot and N. Férey, 2012, Advances in Human-Protein Interaction - Interactive and Immersive Molecular Simulations, Protein-Protein Interactions - Computational and Experimental Tools, Dr. Weibo Cai (Ed.), ISBN: 978-953-51-0397-4, InTech, DOI: 10.5772/36568.

16 A. Saladin, C. Amourda, P. Poulain, N. Férey, M. Baaden, M. Zacharias, O. Delalande and C. Prévost, Nucleic Acids Research, 2010, **38**, 6313.

17 B. Legrand, E. Giudice, A. Nicolas, O. Delalande and E. Le Rumeur, Plos One, 2011, **6**, e23819.

18 A. Roy, A. Kucukural and Y. Zhang, Nature Protocols, 2010, **5**, 725.

19 M. Muthu, K. A. Richardson and A. J. Sutherland-Smith, PLoS One, 2012, **7**, e40066.

20 M. Delarue and Y. H. Sanejouand, J Mol Biol, 2002, **320**, 1011.

21 F. Tama and Y. H. Sanejouand, Protein Engineering, 2001, **14**, 1.

22 S. Schneider, A. Saladin, S. Fiorucci, C. Prévost and M. Zacharias, Methods Mol Biol, 2012, **819**, 221.

23 M. Zacharias, Proteins, 2005, **60**, 252.

24 X. Periole, M. Cavalli, S. J. Marrink and M. Ceruso, J Chem Theor Comp, 2009, **5**, 2531.

25 C. Globisch, V. Krishnamani, M. Deserno and C. Peter, PloS One, 2013, **8**, e60582.

26 D. A. Case, T. E. Cheatham III, T. Darden, H. Gohlke, R. Luo, K. M. Merz Jr., A. Onufriev, C. Simmerling, B. Wang and R. Woods, J Comput Chem, 2005, **26**, 1668.

27 N. A. Baker, D. Sept, S. Joseph, M. J. Holst and J. A. McCammon, Proc Natl Acad Sci U S A, 1998, **98**, 10037.

28 M. V. Petoukhov, D. Franke, A. V. Shkumatov, G. Tira, A. G. Kikhney, M. Gajda, C. Gorba, H. D. T. Mertens, P. V. Konarev and D. I. Svergun, J Appl Cryst, 2012, **45**, 342.

29 W. Humphrey, A. Dalke and K. Schulten, J Mol Graph, 1996, **14**, 33.

30 O. Delalande, N. Férey, G. Grasseau and M. Baaden, J Comput Chem, 2009, **30**, 2375.

31 E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng and T. E. Ferrin, J Comput Chem, 2004, **25**, 1605.

32 T. V. Pyrkov, A. O. Chugunov, N. A. Krylov, D. E. Noldel and R. G. Efremov, Bioinformatics, 2009, **25**, 1201.

33 J. Sarkis, J. F. Hubert, B. Legrand, E. Robert, A. Chéron, J. Jardin, E. Hitti, E. Le Rumeur and V. Vié, J Biol Chem, 2011, **286**, 30481.

34 E. Giudice, A. E. Molza, Y. Laurin, A. Nicolas, E. Le Rumeur and O. Delalande, Biochemistry, 2013, **52**, 7777.

35 K. Y. Chan, J. Gumbart, R. McGreevy, J. M. Watermeyer, B. T. Sewell and K. Schulten, Structure, 2011, **19**, 1211.

36 S. C. Flores, Nucleic Acids Res, 2013, online doi:10.1093/nar/gkt906
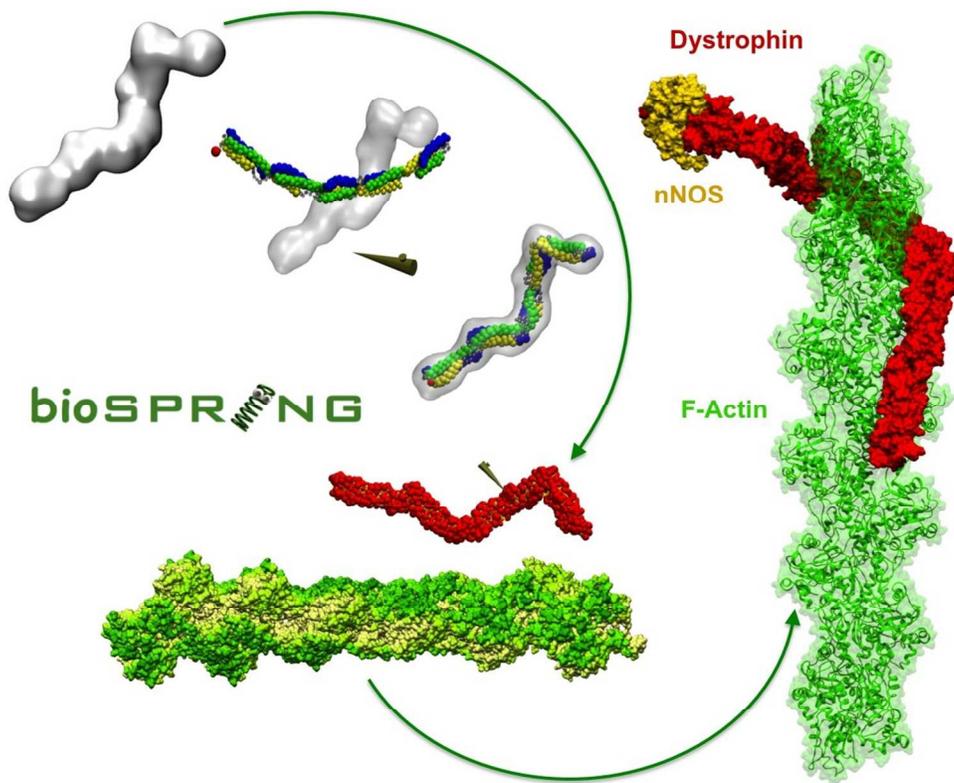
443x362mm (72 x 72 DPI)
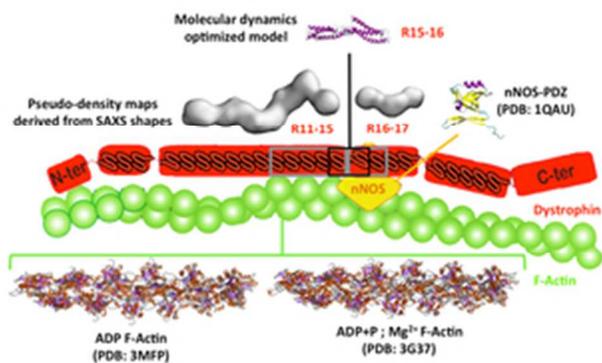
Figure 1: The interaction between dystrophin, nNOS and F-Actin. Structural information for the nNOS-PDZ domain and F-Actin filaments are respectively provided by X-ray diffraction and CryoEM experiments. PDB codes are 1QAU for nNOS and 3MFP or 3G37 for F-Actin. Dystrophin structural data were obtained in solution from Small-angle X-ray Scattering (SAXS) or in silico from homology modelling followed by molecular dynamics optimization.
26x16mm (300 x 300 DPI)
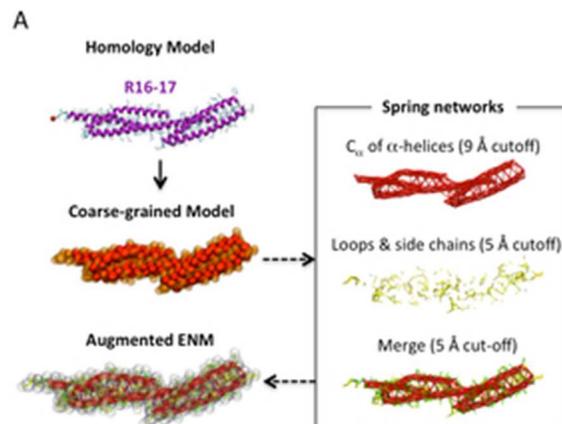
Figure 2A: Protocol for building a coarse-grained augmented Elastic Network Model (aENM) of the R16-17 repeat fragment of dystrophin. From the all-atom homology model provided by the i-TASSER server, a low-resolution model was obtained using the Attract force field parameters developed by Zacharias et al (23). A multi-cut-off spring network can be designed to discriminate between the most flexible and the most rigid regions. Here we chose to stabilize secondary (α-helices) and ternary (coiled-coil) structures. The completion of an aENM model is possible through combining the steric and charge parameters assigned to the pseudo-atoms with the springs defining the ENM.
25x19mm (300 x 300 DPI)
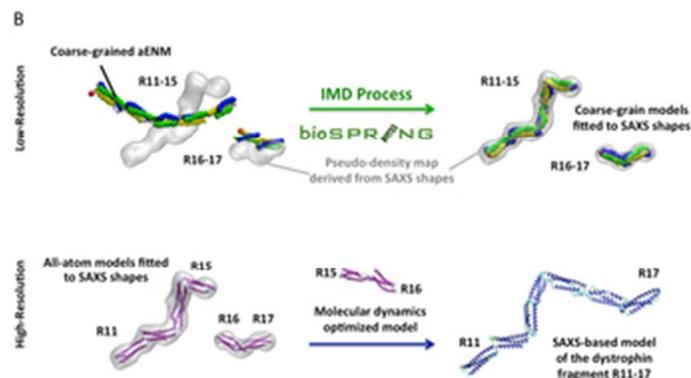
Figure 2B: (Top row) Schematic representation of BioSpring's interactive molecular dynamic processing of the R11-15 and R16-17 dystrophin fragments at low-resolution. For aENM, helices HA are yellow, helices HB are blue, and helices HC are green. The red spot represents the N-terminal end; coils are gray. SAXS shapes were obtained from the GASBOR program of the ATSAS suite then converted to pseudo-density maps for driving the aENM model into the volumetric data. (Bottom row) Fragment reconstruction of an all-atom model is performed by superimposition onto the coarse-grain model then finalized by an optimization into the pseudo-density maps. The full-length filament is obtained by bridging the two SAXS-based models with a molecular dynamics-derived model of the R15-16 fragment, unavailable experimentally.
29x16mm (300 x 300 DPI)
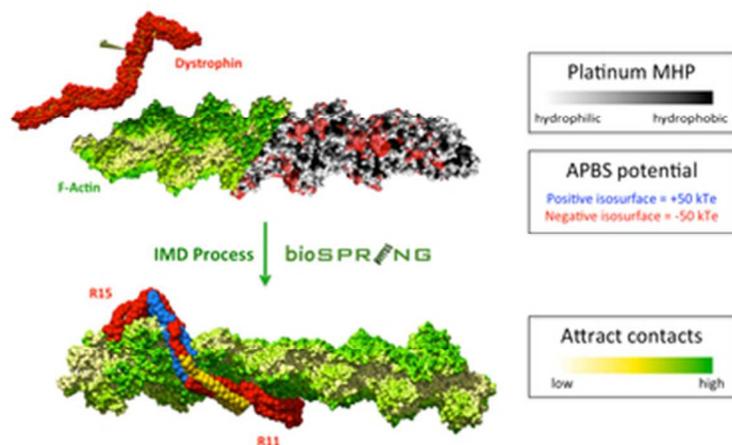
Figure 3: (Top) Starting point for Interactive Flexible Low-Resolution Docking. A dystrophin fragment (R11-15) is modelled by aENM shown here as red van der Waals spheres (coarse grain particles) and yellow sticks (springs). The actin filament's molecular surface (PDB code 3G37) is shown shading either from yellow to green by increasing probability of contact with dystrophin as indicated by Attract docking results, or from white (hydrophilic) to black (hydrophobic) according to Molecular Hydrophobicity Potential (MHP) provided Platinum server (32). The electrostatic potentials calculated using APBS software (27) are shown for +/- 50 kTe isosurfaces and are blue (+) or red (-). These visual representations can be used to guide the user in the interactive experiments. (Bottom) Best pose of the R11-15 fragment after optimization of the aENM model into the potential grid computed for the actin filament structure. In the R11-15 dystrophin regions yellow residues are and blue are not protected against trypsin digestion upon the interaction of the protein with small unilamellar vesicles.
32x19mm (300 x 300 DPI)
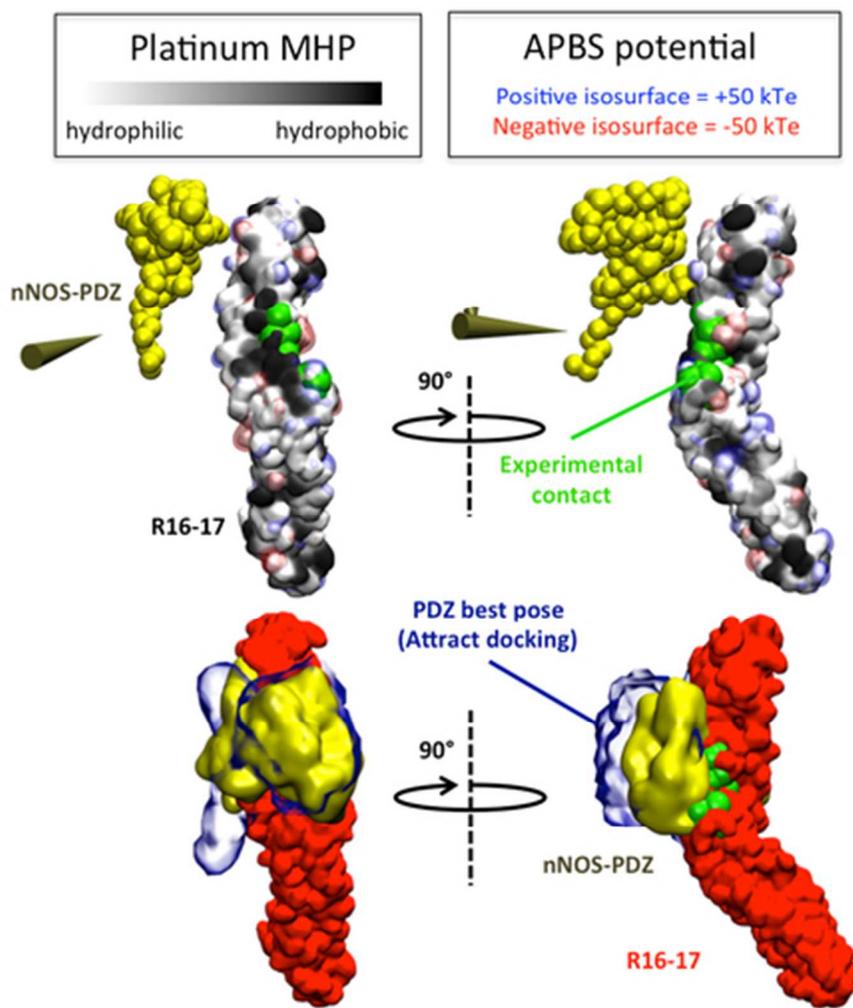
Figure 4: (Top) Starting point for Interactive Flexible Low-Resolution Docking. The dystrophin fragment (repeats R16-17) is shown with a molecular surface shading from white (hydrophilic) to black (hydrophobic) according to Platinum MHP (32). APBS potentials (27) are +/- 50 kTe isosurfaces shown respectively in blue and red. The nNOS-PDZ subdomain is simulated by an aENM in yellow van der Waals spheres. The green spheres highlight the contact regions defined by mutagenesis performed in vitro and in vivo by Lai et al. (Bottom) The nNOS:Dystrophin complex obtained after interactive BioSpring docking, showing the best correlation with the experimental contact mapping. The blue phantom envelope corresponds to one of the best poses, based on complex energy and weighted after clustering, obtained from systematic rigid low-resolution docking (Attract).
40x44mm (300 x 300 DPI)
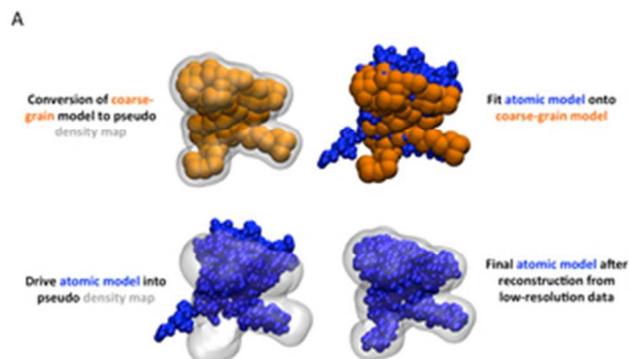
Figure 5A: BioSpring protocol for the atomic reconstruction of the low-resolution model obtained from interactive flexible docking.(Top left) The final position of the coarse-grain aENM in the molecular complex converted into a pseudo density map. (Top right) The atomic aENM is superimposed (Cα RMSD) onto this same coarse-grain aENM final position. (Bottom left) The atomic aENM is driven into the pseudo density map. (Bottom right) The final atomic model is obtained after a few seconds of interactive BioSpring simulation.

26x15mm (300 x 300 DPI)
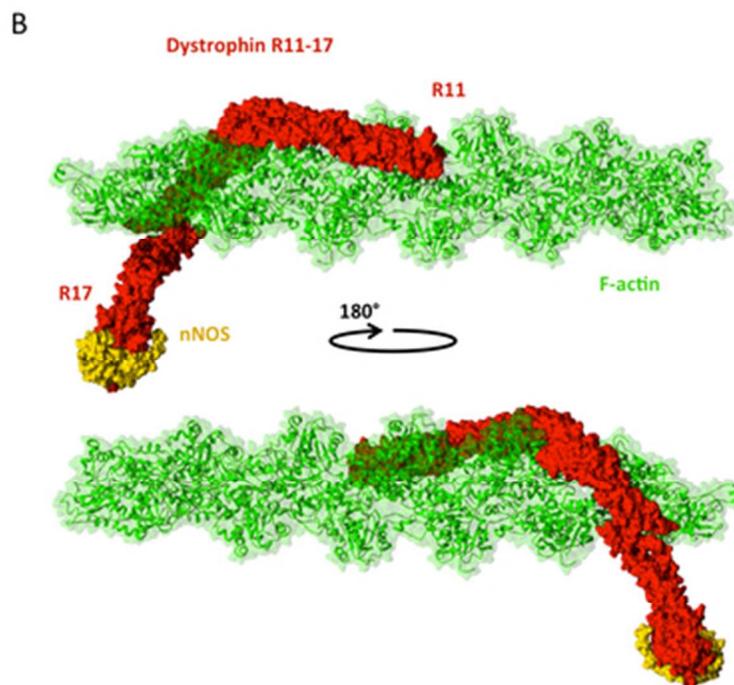
Figure 5B: Final ternary complex obtained by the association of dystrophin to F-actin and nNOS. Dystrophin is red, F-actin (PDB code 3G37) is green, and nNOS is yellow. The complex was developed by combining theoretical and experimental biostructural data with an interactive flexible approach.
34x30mm (300 x 300 DPI)