PROTEINS: Structure, Function, and Bioinformatics 66:87–95 (2007)

Computational Diagnosis of Protein Conformational Diseases: Short Molecular Dynamics Simulations Reveal a Fast Unfolding of r-LDL Mutants That Cause Familial Hypercholesterolemia

S. Cuesta-López,[1,2] F. Falo,[1,2]* and J. Sancho[2,3]*
[1]*Departamento de Física de la Materia Condensada, Universidad de Zaragoza, Spain*
[2]*Instituto de Biocomputación y Física de Sistemas Complejos (BIFI), Universidad de Zaragoza, Spain*
[3]*Departamento de Bioquímica y Biología Molecular y Celular, Universidad de Zaragoza, c/Pedro Cerbuna 12, 50009 Zaragoza, Spain*

ABSTRACT **The molecular basis of conformational diseases frequently resides in mutant proteins constituting a subset of the vast mutational space. While the subtleties of protein structure point to molecular dynamics (MD) techniques as promising tools for an efficient exploration of such a space, the average size of proteins and the time scale of unfolding events make this goal difficult with present computational capabilities. We show here, nevertheless, that an efficient approach is already feasible for modular proteins. Familial hypercholesterolemia (FH) is a conformational disease linked to mutations in the gene encoding the low density lipoprotein receptor. A high percentage of these mutations has been found in the seven small modular binding repeats of the receptor. Taking advantage of its small size, we have performed an in depth MD study of the fifth binding repeat. Fast unfolding dynamics have been observed in the absence of a structural bound calcium ion, which agrees with its reported essential role in the stability of the module. In addition, several mutations detected in FH patients have been analyzed, starting from the native conformation. Our results indicate that in contrast with the wild type protein and an innocuous control mutant, disease-related mutants experience, in short simulation times (2–8 ns), gross departures from the native state that lead to unfolded conformations and, in some cases, to binding site desorganization deriving in calcium release. Computational diagnosis of mutations leading to conformational diseases seems thus feasible, at least for small or modular pathogenic proteins. Proteins 2007;66:87–95.** © 2006 Wiley-Liss, Inc.

Key words: protein unfolding; conformational diseases; r-LDL; familial hypercholesterolemia; molecular dynamics; protein stability; protein folding

INTRODUCTION

Over the last few years, it has become overwhelmingly clear that many diseases are caused by genetic mutations that do not alter protein residues responsible for binding and or catalysis, but simply affect to residues essential for the proper folding and stability of a given polypeptide.[1] Since the number of the latter kind of residues is probably larger than that of those directly located in active sites, it should not be surprising that a majority of genetic diseases are eventually related to conformational defects.[2,3] Protein conformational diseases arise from the fact that the native conformation cease to be the vastly dominant one. They can be divided into two major types depending on whether the native state becomes destabilized relative to the unfolded state or relative to aggregated species. The former type of disease simply leads to a loss of biological function and is more amenable to computational analysis than the latter because the molecular defect appears confined to single copies of the defective polypeptide.

Molecular dynamics (MD) simulations have been quite successful in characterizing protein dynamics, including protein folding[4,5] and, computing efficiency not limiting, it would be possible to simulate and compare the folding of wild type and mutant protein sequences to establish whether a given mutation is compatible with a stable native-like fold or rather leads to a misfolded conformation. However, present computing capabilities can only fold small proteins through very time-consuming computations, which makes the approach unfeasible. The task could be simplified by starting from native folds and following unfolding pathways of mutant sequences, but it would still take to much time given the average size of proteins and the time scale or unfolding processes.

Grant sponsor: Spanish Ministry of Education; Grant numbers: BFM2002-00113, FIS2005-00337 DGES, BFU 2004-01411 MEC; Grant sponsor: Spanish Ministry of Science and Education; Grant number: FPU-AP2002-3492.

*Correspondence to: J. Sancho, Departamento de Bioquímica y Biología Molecular y Celular, Universidad de Zaragoza, c/Pedro Cerbuna 12, 50009 Zaragoza, Spain E-mail: jsancho@unizar.es or F. Falo, Departamento de Física de la Materia Condensada, Universidad de Zaragoza., 50009 Zaragoza, Spain. E-mail: fff@unizar.es

Received 15 May 2006; Accepted 19 July 2006

Published online 16 October 2006 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.21181

Protein functions are typically mediated by the ability to establish specific complexes, which in many cases involve interactions between single domains from each partner. Large proteins, however, use sometimes a different strategy based on the exploitation of a modular nature, as they contain concatenated, homologous domains involved in binding the biological target, that is zinc finger containing proteins or choline binding proteins.[6,7] These binding modules are often quite small and tend to display low secondary structure contents, which is compensated by an abundance of disulfide bridges and bound metal ions. Interestingly, their small size makes them amenable to computer simulation. The low density lipoprotein receptor (r-LDL) is an archetypal modular protein, representative of an entire class of transmembrane receptors[8–10] involved in a large variety of functions such as the clearance of protein complexes from plasma, or the transmission of extracellular signals in nervous system development.[11–14] It consists of seven binding LA modules with a high degree of homology, followed by two EGF-like domains, one β-propeller, one further EGF-like domain, and one glycosilated, one transmembrane, and one cytoplasmic domain.[15] The LDL-r captures circulating LDLs from the plasma and delivers them to the endosome through a complex pH-linked mechanism. The task of binding the LDL particles is performed by the N-terminal LA modules, each one consisting of around 35 residues that present a particular arrangement containing a three disulfide bridge pattern, a complex cooperative hydrogen bond network, and one bound calcium ion. Modules LA4 and LA5 seem crucial for both LDL binding and release. Defects found in any of the r-LDL functional domains can lead to receptor misfunction, deriving in high plasma LDL levels, which is characteristic of familial hypercholesterolemia (FH), a monogenic disorder affecting 1:500 people worldwide. Since the biological basis of FH were found to reside in the LDL receptor,[16] more than 900 different mutations have been reported.[17–20] All these mutations cause the receptor to exhibit different degrees of activity, which cooperates with other factors to produce a variety of clinical manifestations in FH patients. Since different mutations in the same domain can result in different clinical phenotypes, the individual characterization of disease-causing mutations is important for accurate diagnosis, genetic counseling, and a proper treatment.

In this paper we explore the possibility that LA modules exhibit fast enough dynamics so that their native conformation can be used as a starting point to perform MD simulations of disease-related mutants at physiologic temperatures and observe, in short simulation times, unfolding events associated to their expected lower stability. This approach could be used to detect pathologic mutations that are directly related to structural stability and folding and could, in principle, help to investigate other diseases involving small modular proteins. Our results reproduce the experimentally observed stability of the isolated LA5 module when it bears a bound calcium ion, and reveals that the module becomes fully unfolded in the nanosecond time-scale upon removal of the ion, in agreement with the experimental observation of calcium-dependent stability[21] and with our hypothesis of fast unfolding dynamics. Importantly, all the simulations performed on FH-related point mutants lead to severe structural distortions in short simulation times, which strongly suggests that MD simulations of protein modules can easily discriminate disease-related mutations from innocuous ones, making the computational diagnosis of nonaggregative protein conformational diseases a close reality.

MATERIALS AND COMPUTATIONAL METHODS

The crystal structure of the LA5 module was extracted from the Protein Data Bank (pdb code 1ajj).[22] The different mutations in the module were initially designed in silico using the Swiss-Pdb Viewer mutation tool[23] and, previously to the simulation process, refined through a minimization scheme that uses several cycles of Steepest Descent, Conjugate Gradient, and Adopted Basis Newton-Raphson algorithms up to 20,000 steps in total.

MD simulations were performed using the academic version of the CHARMM (c27b2) package.[24] Solvation of the systems was achieved placing a pre-equilibrated cubic box ($50 \times 50 \times 50$ Å) of ~6000 TIP3P water molecules[25] around all protein structures. Periodic boundary conditions have been applied using the CHARMM *Crystal* facility to reduce edge effects, and the SHAKE algorithm[26] has been used to hold rigid the internal geometry of the water molecules (Jorgensen description). In addition, long-range electrostatic interactions were modeled with the particle-mesh Ewald method,[27] using a cut-off of 12.0 Å and a grid space between 0.95 and 1.0 Å. Solvent environment set up was complemented with a placement of Na^+ counterions to reach an appropriate neutralization. Diffusion of the counterions and suppression of the potential internal strains present in the solvation cage were achieved through 100 ps (300 K, 1 at) of CPT dynamics[28] with a fixed solute. In the special case of the calcium removal simulation, an additional neutralization and equilibration of the solute environment was required. After a slow progressive heating of 50 ps to the working temperature (310 K) and a proper equilibration phase (50–200 ps), different techniques such as Brownian Dynamics or a Nosé-Hoover thermostat were applied for Canonical trajectories production. Main events were preserved, although with different time scale dynamics, giving a global consistence to the computer experiment. The friction coefficient γ in the Langevin equations was set to 64 ps^{-1} for the solvent molecules, while a much lower one (0.0–1.5 ps^{-1}) was used for the atoms in the protein and closer solvent molecules in a radial sphere of 25 Å principal component analysis[29–31] was performed with user made subroutines in cooperation with the VIBRAN module implemented in CHARMM. Solvent accessible surface area (SASA) calculations have been performed using Naccess 2.1.1,[32] in connexion with CHARMM through a home-made program based on the first method proposed by Lee and Richards,[33] using a probe sphere of 1.4 Å. Molecular visual graphics were generated using Swiss-Pdb Viewer.[23]

RESULTS

Ca^{2+} Dependent Integrity of the LA5 Module, and Fast Unfolding Kinetics

In the presence of calcium, the LA5 repeat adopts a polypeptide fold common to all LA modules [Fig. 1(a)]. It is stabilized by three disulfide bonds and a bound calcium ion. The pattern of disulfide bonds and the calcium binding sequence are conserved in the seven repeats of the LDL receptor. The LA5 module is made of two "lobes". The N-terminal one comprises a short tail linked by a disulfide bond to a β-hairpin. The C-terminal lobe includes a further disulfide bond and the calcium site, where the ion is trapped in an octahedral coordination cage [Fig. 1(b)] formed by the acidic side chains of Asp196, Asp200, Asp206, and Glu207 (Asp25, Asp29, Asp35, and Glu36 according to 1ajj pdb-file numbering) and the backbone carbonyls of Trp193 and Gly198 [Trp22 and Gly27 in Fig. 1(b)].[22] There is a small hydrophobic core between the lobes where the conserved Phe181 and Ile189 side chains pack onto apolar atoms of the calcium binding residues. The lobes are further tightened together by one disulfide bridge and, additionally, they establish a complex hydrogen-bonding network that includes Glu187-Lys202 and Ser185-Asp203 interactions. Nuclear magnetic resonance studies of wild-type LA5 and of the LA5-LA6 unit have demonstrated an essential role of the calcium ion in maintaining the structural integrity of the folded structure.[21,34] In addition, it has been shown[35] that calcium is needed in vitro to guide the formation of native disulfide bonds.

We have thus performed MD simulations of the wild type LA5 module (in the presence and absence of calcium) to probe the structural role of the ion and to estimate the time scale of LA5 unfolding events associated to destabilization of the native structure. After a technical process of solvation and equilibration at physiological temperature (310 K, see Methods), the bound cation has been removed and the subsequent relaxation process has been observed. According to our Canonical MD trajectories, calcium removal triggers a fast sequence of events, leading to a severely unfolded state. The increment of SASA and the mean-square fluctuations of the system along the unfolding pathway are shown in Figures 2 and 3, respectively.

The unfolding is initiated by changes in the coordinating sphere. The first unfolding event starts very early, at around 0.5 ns, when Asp196 turns its side chain to the solvent to minimize repulsions with the neighboring Glu207. This is followed (MD time 1.0 ns) by a similar evolution toward full solvent exposure of the hydrophobic Trp193 and the acidic Glu207 residues (see Fig. 3). The global unfolding starts at around 1.5 ns and leads to a complete disorganization of the calcium binding site at 2.0 ns. Clearly, the structural effects associated to calcium removal extend beyond the binding site and give rise to a severe disruption of the interactions between the two lobes of the module. In the absence of the cation, Lys202 and Asp203 display high conformational flexibility and the hydrogen bonds with their native partners at the N-terminal lobe get broken. Note, for instance (see Fig. 3), how soon Ser185 and Glu187 also become exposed after Asp203 gets in contact to surrounding water molecules. Similarly, the small hydrophobic core between the lobes is disrupted and Ile189 becomes fully exposed. The final metastable state reached (last pic inside Fig. 2) is probably as unfolded as a 37-residue polypeptide with the intact disulfide pattern of the LA modules can be. In contrast (see Fig. 2), when the wild-type module is simulated with its bound calcium ion under, otherwise, identical conditions, it remains perfectly stable and preserves all the structural elements characteristic of the native LA5

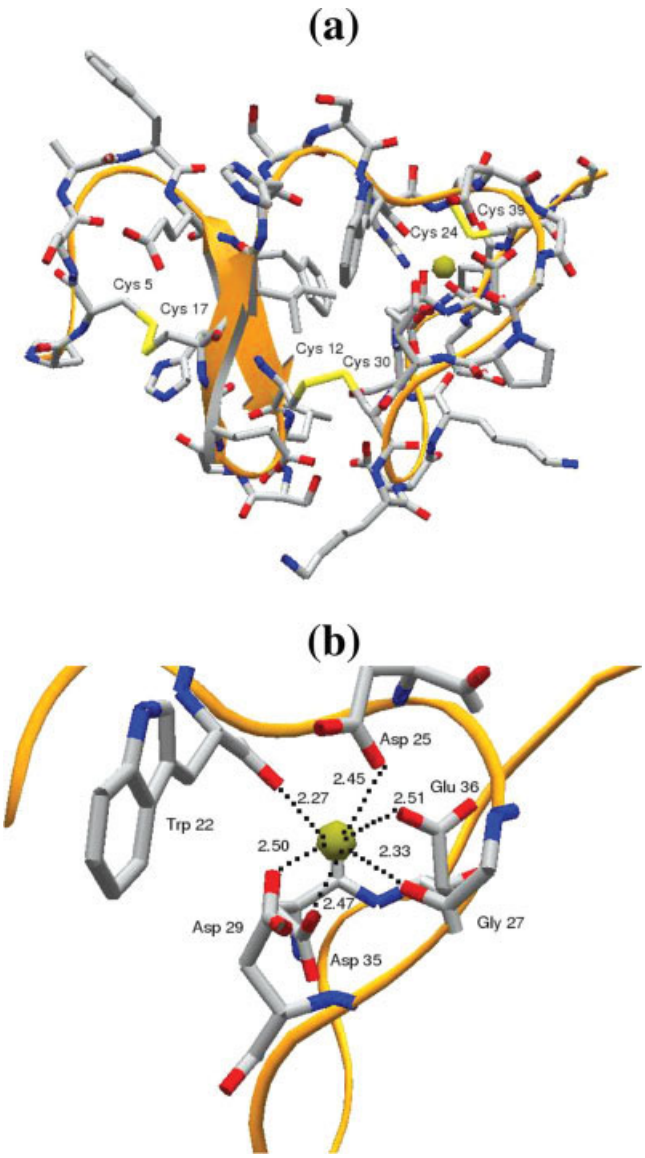

Fig. 1. (**a**) X-ray structure of the LA5 module (pdb-code 1ajj), and ribbon diagram showing a small β-hairpin in the N-terminal lobe and the Ca^{2+} binding region in the C-terminal lobe. The three disulphide bridges are shown in yellow. (**b**) Zoom of the octahedral calcium binding site described by two backbone carbonyl groups and four carboxyl side chains. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

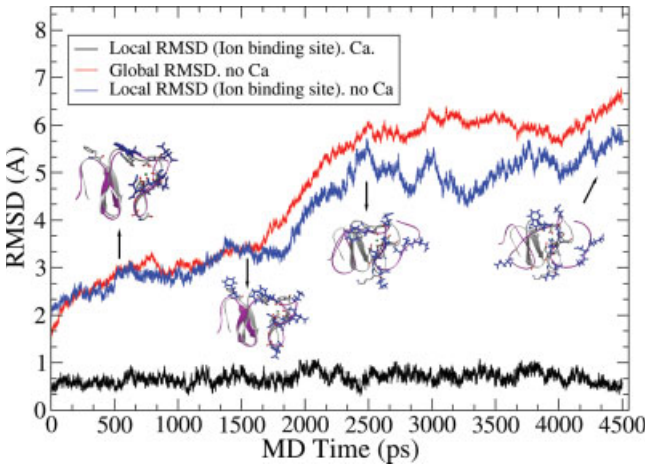

Fig. 2. RMS deviation of the LA5 module, relative to the crystal structure, after calcium removal. A large departure from the crystal structure, leading to severe unfolding, is triggered by removal of the bound ion from the wild type module. In contrast, when the calcium ion is present, the binding site is preserved. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

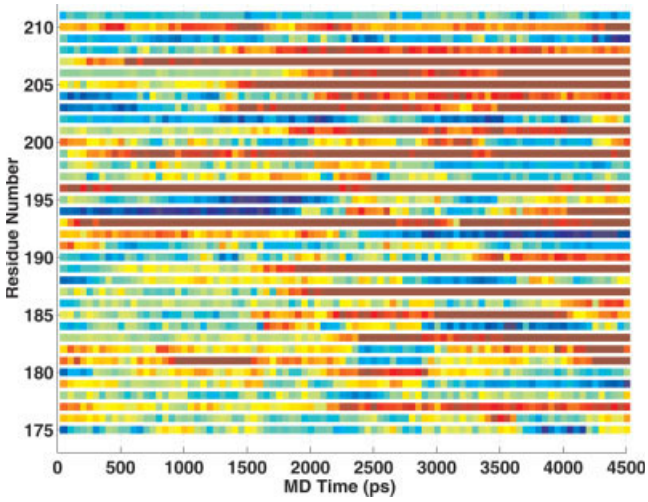

Fig. 3. Deviation in SASA per residue along a 4.5 ns trajectory of Brownian Dynamics recorded after calcium removal. Changes are represented in a color scale that runs from deep blue -50 A^2 to hot red $+50$ A^2, describing burial or exposure, respectively. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

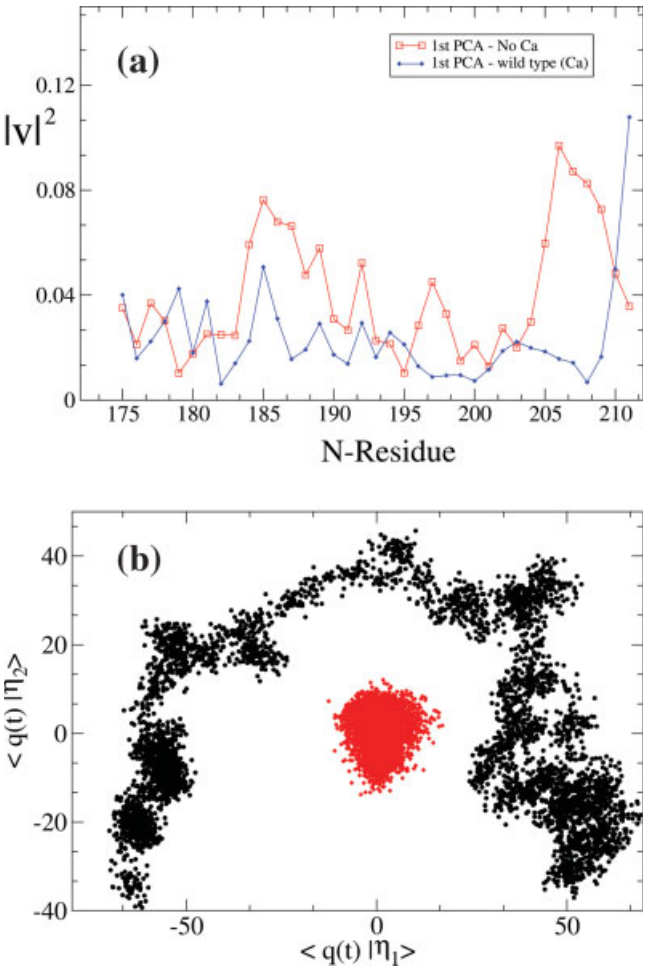

Fig. 4. Square amplitude of the first principal component mode per residue for different MD realizations. The unfolding curve for the LA5 module in the absence of calcium (red lines) shows large fluctuations near the hydrophobic core. The two main peaks in the 1st component represent a mode that pushes apart the *lobes* of the module, and exposes the hydrophobic core and the calcium cage. It is thus a clear unfolding conformational change. In contrast, the native module remains rather stable, showing only small fluctuations. (Bottom) Plane described by the projection over the MD trajectory of the first two principal component modes. The simulation in the absence of calcium (black dots) describes an evolution from the folded to the unfolded states. The central cluster (red dots) shows the stability of the native structure. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

structure (i.e. the octahedral calcium binding cage, the hydrogen bonding network, and the hydrophobic core) until at least 7 ns (time of the simulation, not shown).

These structural effects of the calcium ion have been corroborated by a principal component analysis (PCA) of the trajectories.[29–31] The square amplitude of the principal modes per protein residue for trajectories corresponding both to the LA5 module in the absence of calcium and to the native structure are compared in Figure 4(a). This representation allows a detailed identification of the protein regions that contribute most to the dynamics. Note that the residues forming the octahedral cage display a rather stable conformation in the native module (the amplitudes for residues Asp206, Glu207, Trp 193, and specially Asp200, Gly198, Asp196 show low values when compared with those exhibited in the absence of calcium) and only small fluctuations of the terminal tails and first lobe are observed. In contrast, after calcium removal, a global unfolding event is described by the 1st PC that shows a coupled mode pulling apart the two *lobes* of the protein, thus exposing the hydrophobic core and disorganizing the binding site. In Figure 4(b), the two dimensional plane defined by projections of the first two principal modes is shown. This plane represents a significant percentage of the relevant subspace that contributes to the total mean square fluctuations.[31] The native struc-

TABLE I. Summary of Different Degrees of Disruption (RMSD) and of Criteria for the Evaluation of the Folding State of the Main Structural Motifs in the Wild Type and Mutant LA5 Module (Two Different Simulations for Wild Type *WT*$_{1;2}$)

Protein variant	N-terminal tail (res. 4 : 9) (Å)	β-Hairpin turn (res. 10:19) (Å)	Binding site (Å)	Interlobes distance (Å)	H-bond network	Tot. SASA (Å^2)	SASA H.C. (Å^2)	LDL-r activity (%)	MD time (ns)
D206G	3.52	4.46	4.80	5.8	Not formed	3758.9	220.2	—	5.5
D206E	3.75	5.20	6.05	14.04	Not formed	3796.4	211.14	<2[a]	8.0
	4.21	4.63	3.51	6.94	Not formed	3819.3	217.93		6.5
	1.91	3.08	1.92	9.70	Not formed	3442.2	128.56		5.0
D200Y	2.76	3.20	2.44	10.22	Not formed	3413.1	134.61	<2–8[a]	6.5
D203G	3.02	2.21	1.31	6.8	Not formed	3033.5	79.92	5–15	5.0
C176R	9.17	4.23	3.67	11.80	Not formed	3216.9	107.66	20	5.0
N209A	2.6	1.84	1.85	7.1	Formed	2822.4	124.48	—	6.8
WT$_1$	2.49	1.4	1.12	7.0	Formed	2915.5	59.26	—	8.0
	2.7	1.3	1.23	6.4	Formed	2910.2	55.75	—	6.0
WT$_2$	1.84	1.34	0.95	7.2	Formed	2902.0	83.24	—	5.0
WT cryst.	—	—	—	6.26	Formed	2746.11	61.05	—	—

Main criteria for identification of pathogenic mutations are binding site distortion, H-bond network disruption, and total exposure to the solvent.
[a]Activities for mutations in these sites have been predicted to be less than 2%, although no particular data have been reported for these particular variants.

ture of the protein (red dots) describes a well localized cluster in the two first principal component plane, which means that, when calcium is present, the protein is stable and only small fluctuations are observed. On the contrary, the simulation in the absence of calcium (black dots) describes an evolution from the folded to the unfolded states.

Fast In Silico Unfolding of LA5 Modules Bearing Mutations That Cause FH

The LA5 module is the key for both LDL uptake and their subsequent release at the acidic pH of the endosome.[36] In addition, LA5 is essential for VLDL uptake. The important role of the module in LDL transport is reflected in the large number of known mutations leading to FH that are clustered in LA5. A structural mapping of those mutations indicates that most of them affect to residues involved in some of the structural features that stabilize the native conformation, such as the calcium cage, the disulfide bridges, or the interface of the two lobes.[37] It is thus expected that those mutant modules will be significantly destabilized and that some of them might not even fold into a native-like conformation. In vitro studies have demonstrated that isolated LA5 modules bearing some of the reported mutations are, indeed, unable to develop a correct pattern of disulfide bonds and/or cannot bind calcium with high affinity.[38] The fast unfolding dynamics observed in the simulation of the isolated, Ca^{2+}-deprived LA5 module (see earlier) has prompted us to investigate whether the lost of competence for native folding of a given mutant LA5 module can be diagnosed in a reverse manner by performing short MD simulations of mutant modules, starting from the native conformation. We have thus selected several pathogenic mutations involving residues located in different structurally important LA5 regions [Fig. 5].[22] The mutant proteins have been simulated and different criteria, summarized in Table I, have been applied to categorize and discriminate disease-related mutations.

Two of the mutations take place in the conserved acidic sequence involved in Ca^{2+} coordination: D200Y and D206E. The first one, D200Y, has been described among others in Finnish and German population.[39] The wild type calcium coordinating Asp200 side chain seems to play an important role in maintaining protein functionality because the rather conservative D200N mutation is also known to lead to FH, and D200G has been reported to decrease LDL-r function to 8%.[20] MD simulations of D200Y show that, initially, the bulkier uncharged tyrosine residue makes the calcium ion come slightly closer to the Asp196 and Glu207 coordinating residues, and forces an expansion of the binding site. After 5 ns, the expansion evolves so that several residues located in the C-terminal lobe, Asp203 and Ser205 (see Fig. 6), become exposed to solvent and then, the Glu187-Lys202 interaction between the two lobes of the module is broken, these residues also becoming exposed. The disruption of the hydrogen bonding network between the lobes is likely to be followed by further unfolding, not observed in the first 5 ns. A comparison is shown in Figure 6 of the conformations exhibited by the metastable states of this and other mutant LA5 modules, after 5 ns of MD, with the original wt-crystal structure in terms of changes in SASA for each residue. Additionally, RMS deviations for characteristic substructures of the module are presented in Table I.

The second mutation studied at the calcium cage, D206E, is characteristic of Afrikaner population[17] and involves the substitution of the coordinating aspartate by the slightly longer, but equally charged, glutamate. This mutation similarly leads to a clear unfolding, which is a

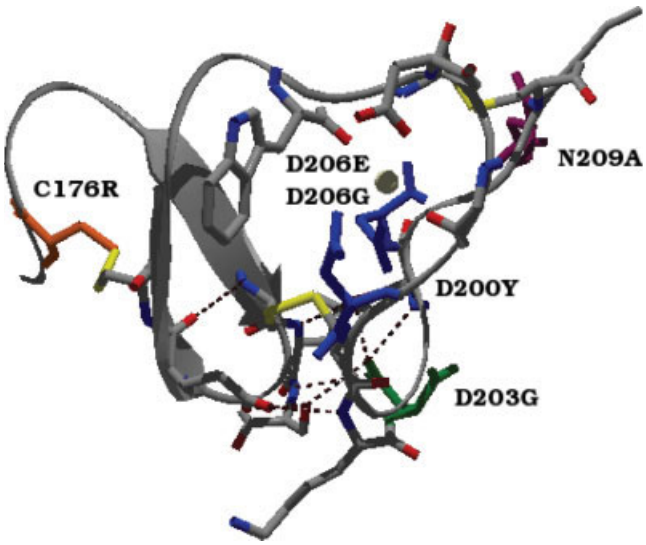

Fig. 5. Location in the LA5 module of residues critical for folding and corresponding mutations selected for computational study. Blue residues are mutations that affect directly the calcium binding site, the green residue is involved in the H-bond network between the two lobes of the module, and the orange one breaks the N-terminal disulfide bridge. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

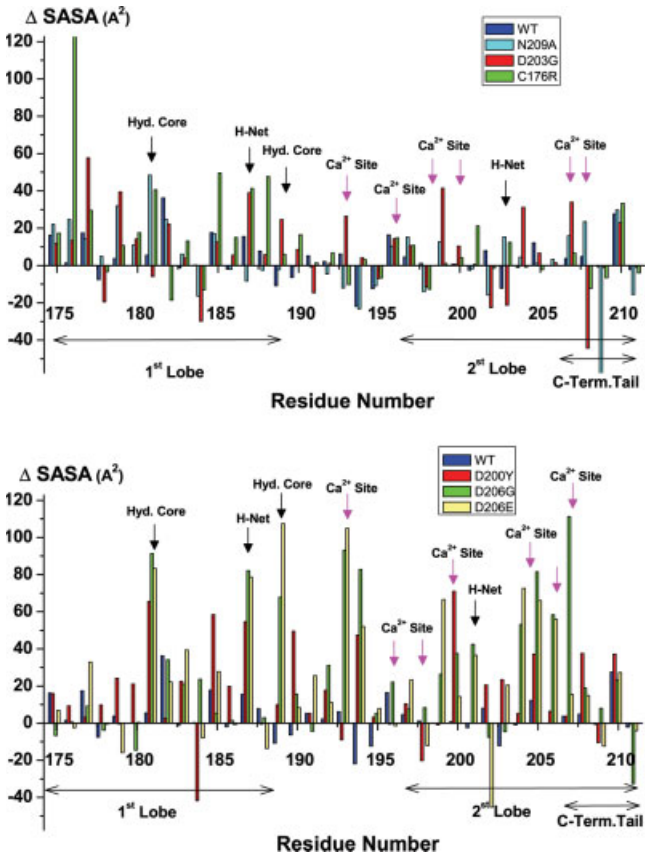

Fig. 6. Comparison for each variant simulated of increments in SASA, at the residue level, in stable states obtained after 5 ns of MD. The crystal structure of the native conformation has been taken as reference. Note that FH-related mutations cause relevant changes of exposure in the main structural components of the module. Left graph represents mutations at the binding site, causing global unfolding phenomena manifested in structural motifs such as the hydrophobic core (Hyd.Core) or the H-bond network (H-net). Right graph compares pathogenic replacements involving the disulphide pattern (C176R) or the H-bond Network (D203G) with an innocuous substitutions (N209A) and with WT. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

bit slower. At 6.5 ns, the calcium cage is disorganized and the coordinating residues are exposed to solvent, some making hydrogen bonds with water molecules. Data in Table I confirm that a fully unfolded structure is reached in 8 ns, which is consistent with reports by Blacklow and Kim,[21] showing that this mutant is unable to fold properly in the presence of calcium and suggesting that the structural instability of the module is coupled to the distortion of the calcium cage. It seems clear from the present simulations that efficient calcium coordination requires the integrity and precise geometry of the calcium cage. In addition, the stability of the Ca^{2+} building site appears essential to preserve the interlobe H-bonding network, which is needed to keep the required solvent accessibility for residues His190, Ser191, and Lys202, which form, in cooperation with the conserved acidic residues, the interface that interacts with the β-propeller in the LDL release model.[40]

The third mutation C176R has been found in FH patients in the United Kingdom,[41] and leads to 20% LDL receptor activity detected in cells, relative to normolipaemic controls. Other mutations at the same position (C176F and C176Y) have been reported to lower the activity below 2%.[17,19] The replacement of cysteine 176 by arginine breaks the 176–188 disulfide bond present in the wild type structure. This bond appears to maintain the N-terminal tail of the LA5 module packed against the β-hairpin, thus contributing to the stability of the N-terminal lobe. According to our simulation, the first consequence of the C176R mutation is a significant increment in conformational entropy of the N-terminal tail, which allows it to explore conformations that were previously restricted by the disulfide bond. The greater flexibility of the tail seems to destabilize the packing between the two lobes, which leads to a partial exposure of the hydrophobic core [significantly that of Phe181, see Figs. 6 and 7(c)]. In addition, the fluctuations are transferred to the hydrogen bonding network connecting the two lobes, which is debilitated so that the lobes become disengaged and their relative distance largely increased (Table I). In spite of the destabilization of the N-terminal lobe and of its interactions with the C-terminal one, the latter retains a close to native conformation at the calcium binding site, where the ion remains bound. This retention of calcium binding capability in the absence of the 176–188 disulfide bond has been experimentally observed by Koduri and Blacklow.[42] Since the mutant module retains its calcium binding site and presents a structure close to the native conformation for the C-terminal lobe, we propose that the link between the mutation and FH may be found in an increase of the average distance between modules LA4 and LA5. The precise docking of these modules on the β-propeller domain

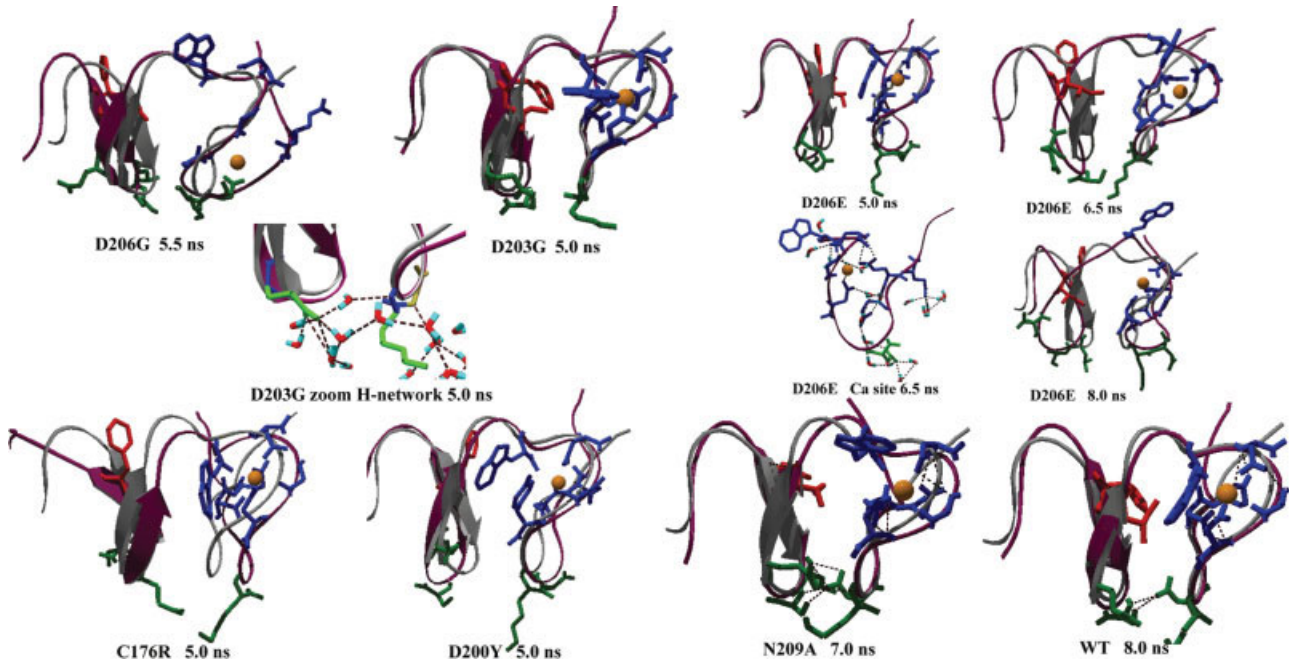

Fig. 7. Conformational states reached by the different variants. A ribbon diagram in white corresponding to the crystal structure is always presented as a comparison reference. Residues involved in calcium coordination are colored in blue, and hydrophobic core residues in red. The calcium cage is severely distorted in all the substitutions of the coordinating residues. Note that the H-bond network only remains stable in the wild-type and N209A proteins (residues colored in green). Unfolding evolution from 5–8 ns is presented for mutation D206E. Changes in the binding region for this protein lead to a totally unfolded module starting with a hydration of the calcium site (D206E 6.5 ns). In the case of mutation D203G the H-bond network is not stabilized and a highly hydrated surface appears instead.

seems essential for LDL release at endosomal pH and a lengthening of the intermodule distance could be detrimental.

The fourth pathogenic mutation simulated is D203G, which is widely found in Italian and American individuals.[17] This mutation has been reported to decrease the receptor activity to 5–15%, and other mutations at the same position (D203N and D203A) have also been reported to cause FH. From the structural point of view, the D203G mutation suppresses an inter-lobe hydrogen bond between Asp203 and Ser185, and one hydrogen bond between Asp203 and Ser205, which helps to position Ser205 to bind Leu184. The D203G mutation could thus disrupt the hydrogen bonding network connecting the two lobes of the LA module. The simulation clearly shows that the network is disrupted. In the absence of the Asp203 side chain, the link to the backbone amide of Ser185 is no longer formed, which distorts the turn involving the backbone amide of Ser205, close to the binding site. Besides, the inter-lobe bond formed between Glu187 and Lys202 disappears. The perturbation is propagated so that residues close to the calcium cage (such as Pro199) increase their solvent exposure, and even Trp193 and Glu207 are able to explore exposed conformations along the trajectory. A consequence of this debilitation of the calcium cage associated to the D203G mutation has been experimentally observed by Blacklow and Kim,[21] whose studies of folding calcium dependency showed the mutant is only able to refold in the presence of a calcium concentration 20 times higher than that required by wild-type LA5.

DISCUSSION

The LA5 module and its mutants display fast conformational dynamics. This is clearly shown by the fast global unfolding of the module (in <5 ns) after calcium removal, which is consistent with previously reported experiments, indicating an essential role of calcium for the integrity of the LA5 module.[21,34] Taking advantage of the fast kinetics observed, we have explored the effect exerted by several single point mutations in the structural stability of the LA5 module. Our results indicate that short MD simulations can be used to distinguish disease-related FH mutations from innocuous ones. The structural departures from the native state observed in the simulations of the mutants are shown in Table I. All mutants involved in FH present a disorganized inter-lobe H-bond network and significant distortions of both the N-terminal lobe and the calcium binding site, and most show an increased average distance between the two lobes. The global SASA increments observed in the mutants are indicative of unfolding phenomena involving the hydrophobic core. Thus it seems clear that short MD simulations of FH-related mutations in the LA5 module of the LDL-r can reveal unfolding events linked to a reduced structural stability that very likely causes LDL receptor misfunction.

To further assess the performance of the simulations, we have selected two additional mutations, one representative of structurally conservative mutations, unlikely to give rise to LA5 unfolding, and one representative of severely disruptive mutations. In an attempt to identify

innocuous mutations, the LDL polymorphisms database has been searched,[20] but it contains no examples of polymorphisms located in the LA5 module. On the other hand, a search in the chimpanzee genome indicates that its LA5 module is identical to the human one. We have thus designed and simulated the conservative mutation N209A. The wild type Asn209 residue is located in the C-terminal lobe of the module, in a solvent exposed position, and its side chain does not form hydrogen bonds. The simulation of the N209A mutant indicates that the structure experiences, at around 2 ns, small rearrangements in the neighborhood of the mutated residue, otherwise remaining stable until the end of the trajectory (5 ns). Importantly, the hydrogen bond network connecting the two lobes remains stable and the calcium cage perfectly formed. As a representative of a disrupting mutation, we have designed and studied the D206G mutant. Replacement of the calcium coordinating aspartate by a glycine is expected to be more disruptive that its replacement by a glutamate. In this respect, faster unfolding kinetics than those observed for the FH D206E mutation, previously discussed, would be expected. This is indeed the case, and the relaxation path followed by D206G is faster and leads to a more distorted structure at 5 ns (see Fig. 6).

CONCLUSIONS

Simple MD simulations in the nano scale confirm the strong dependence of the conformational stability of the 37-residue LA5 module of the LDL-r on the integrity of specific structural elements (i.e. disulfide bonds, bound ions, H-bond networks). Consistent with previously reported experiments,[21,34] the absence of calcium in the LA5 module leads to quick unfolding. In addition, a fast dynamical response to destabilizing mutations is observed in the module.

Fast (5–10 ns) MD simulations of mutant LA5 modules found in individuals suffering from FH reveal substantial departures of the native conformation, while a control mutation remains stable in the same time scale. This sensibility of the module to mutational perturbations raises the possibility of a computational analysis of not yet-observed single point mutations to anticipate their potential involvement in LDL-r misfunction. We think that other LA modules and, possible, other small modules present in different human proteins could similarly display fast dynamics and might also be investigated with the same purpose and methodology, thus making possible the computational diagnosis of conformational diseases involving small or modular human proteins.

ACKNOWLEDGMENTS

We specially thank Dr. Yves-Henri Sanejouand at ENS-Lyon (France) for his useful comments on protein simulation. We acknowledge computational support and technical advising from Dr. P.J. Martinez.

REFERENCES

1. Dobson C. Protein folding and misfolding. Nature 2003;426:884–890.
2. Wang Z, Moult J. SNPs. Protein Structure, and Disease. Hum Mutat 2001;17:263–270.
3. Yue P, Li ZL, Moult J. Loss of protein structure stability as a major causative factor in monogenic disease. J Mol Biol 2005; 353:459–473.
4. Karplus M, Weaver D. Protein folding dynamics. Nature 1976; 260:404–406.
5. Mayor U, Guydosh N, Johnson C, Grossmann J, Sato S, Jas G, Freund S, Alonso D, Daggett V, Fersht A. The complete folding pathway of a protein from nanoseconds to microseconds. Nature 2003;421:863.
6. Miller J, McLachlan A, Klug A. Repetitive zinc-bending domains in the protein transcription factor IIIA from Xenopus oocytes. EMBO J 1985;4:1609–1614.
7. Sanchez-Puelles J, Sanz J, Garcia J, Garcia E. Cloning and expression of gene fragments encoding the choline-binding domain of pneumococcal murein hydrolases. Gene 1990;89: 69–75.
8. Russell D, Schneider W, Yamamoto T, Luskey K, Brown M, Goldstein J. Domain map of the LDL receptor: sequence homology with the epidermal growth factor precursor. Cell 1984;37:577–585.
9. Yamamoto T, Davis CG, Schneider W, Casey M, Goldstein J, Russell D. The human LDL receptor: a cysteine-rich protein with multiple Alu sequences in its mRNA. Cell 1984;39:27–38.
10. Sudhof TC, Russell D, Goldstein J, Brown M. The LDL receptor gene: a mosaic of exons shared with different proteins. Science 1985;228:893–895.
11. Takahashi S, Kawarabayasi Y, Nakai T, Sakai J, Yamamoto T. Rabbit very low density lipoprotein receptor: a low density lipoprotein receptor-like protein with distinct ligand specificity. Proc Natl Acad Sci USA 1992;89:9252–9256.
12. Saito A, Pietromonaco S, Loo A, Farquhar M. Complete cloning and sequencing of rat gp330/"megalin," a distinctive member of the low density lipoprotein receptor gene family. Proc Natl Acad Sci USA 1994;91:9725–9729.
13. Herz J. The LDL receptor gene family: (un)expected signal transducers in the brain. Neuron 2001;29:571–581.
14. Pinson K, Brennan J, Monkley S, Avery AJ, Skarnes WC. An LDL- receptor-related protein mediates Wnt signalling in mice. Nature 2000;407:535–538.
15. Herz J. Deconstructing the LDL receptor—a rhapsody in pieces. Nat Struct Biol 2001;8:476–478.
16. Brown M, Goldstein J. Familial hypercholesterolemia: a genetic defect in the low-density lipoprotein receptor. N Engl J Med 1976;294:1386–1390.
17. Hobbs HH, Brown MS, Goldstein JL. Molecular genetics of the LDL receptor gene in familial hypercholesterolemia. Hum Mutat 1992;1:445–446.
18. Goldstein JL, Hobbs HH, Brown MS. Metabolic and molecular bases of inherited disease, Chapter 120. New York: McGraw-Hill; 2001.
19. Varret M, Rabes J-P, Viller L, Junien C, Beroud C, Boileau C. The human LDL receptor gene database: molecular analysis of 535 mutations. (Available at: http://www.umd.necker.fr, INSERM U383, Hopital Necker, 75015 Paris, France, 2001).
20. Available at: http://www.ucl.ac.uk/fh/. The low density lipoprotein receptor (LDLR) gene in familial hypercholesterolemia. Mutation database (2001).
21. Blacklow SC, Kim P. Protein folding and calcium binding defects arising from familial hypercholesterolemia mutations of the LDL receptor. Nat Struct Biol 1996;3:758–762.
22. Fass D, Blacklow SC, Kim P, Berger J. Structure of LDL receptor module reveals molecular mechanisms of familial hypercholesterolemia. Nature 1997;388:691–693.
23. Guex N, Peitsch M. Swiss-model and the Swiss-Pdb Viewer: an environment for comparative protein modeling. Electrophoresis 1983;18:2714–2723.
24. Brooks BR, Bruccoleri RE, Olafson BD, Status DJ, Swaminathan S, Karplus MJ. CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. J Comput Chem 1983; 4:187–217.

25. Jorgensen WL, Chandrasehkar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. J Chem Phys 1983;79:926–935.
26. Ryckaert JP, Ciccotti G, Berendsen HJC. Numerical integration of the cartesian equation of motion of a system with constraints: molecular dynamics of N-alkanes. J Comput Phys 1997;23:327–341.
27. Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen L. A smooth particle mesh Ewald method. J Chem Phys 1995;103: 8577–8597.
28. Feller SE, Zhang Y, Pastor RW, Brooks BR. Constant pressure molecular dynamics simulation: the Langevin piston method. J Chem Phys 1995;103:4613–4621.
29. Go N, Noguti T, Nishikawa T. Dynamics of a small globular protein in terms of low-frequency vibrational modes. Proc Natl Acad Sci USA 1983;80:3696–3700.
30. Amadei A, Linssen A, Berendsen H. Essential dynamics of proteins. Proteins: Struct Funct Genet 1993;17:412–425.
31. Hayward S, Go N. Collective variable description of native protein dynamics. Annu Rev Phys Chem 1995;46:223–250.
32. Hubbard SJ, Thornton JM. 1993. NACCESS, Computer Program, Department of Biochemistry and Molecular Biology, University College London, London.
33. Lee B, Richards FM. The interpretation of protein structures: estimation of static accessibility. J Mol Biol 1971;55:379–400.
34. Atkins A, Brereton I, Kroon P, Lee H, Smith R. Calcium is essential for the structural integrity of the cysteine-rich, ligand-binding repeat of the low-density lipoprotein receptor. Biochemistry 1998;37:1662–1670.
35. North CL, Blacklow SC. Structural independence of ligand-binding modules five and six of the LDL receptor. Biochemistry 1999; 38:3926–3935.
36. Rudenko G, Henry L, Henderson K, Ichtchenko K, Brown M, Goldstein J, Deisenhofer J. Structure of the LDL receptor extracellular domain at endosomal pH. Science 2002;298:2353–2358.
37. Machicado C, Bueno M, Castillo S, Pocovi M, Sancho J. Clinical mutations in the r-LDL that group FH with the protein folding diseases. Invest Arterioscl 2003;15:59–65.
38. North C, Blacklow S. Solution structure of the sixth LDL-A module of the LDL receptor. Biochemistry 2000;39:2564–2571.
39. Geisel J, Holzem G, Oette K. Screening for mutations in exon 4 of the LDL receptor gene in German population with severe hypercholesterolemia. Hum Mutat 2003;96:301–304.
40. Jeon H, Blacklow SC. An intramolecular Spin of the LDL receptor β-Propeller. Nature Struct Biol 2003;11:133–136.
41. Webb J, Sun X, McCarthy S, Neuwirth C, Thompson G, Knight B, Soutar A. Characterization of mutations in the low density lipoprotein (LDL)-receptor gene in patients with homozygous familial hypercholesterolemia, and frequency of these mutations in FH patients in the United Kingdom. J Lipid Res 1996; 37:368–391.
42. Koduri V, Blacklow S. Folding determinants of LDL receptor type A modules. Biochemistry 2001;40:12801–12807.