Theor Chem Acc (1999) 101:82–86
DOI 10.1007/s002149800m53

Theoretical
Chemistry Accounts

Regular article

Molecular dynamics simulations of the isolated domain 1 of annexin I*

Tru Huynh[1], Gabriel Musat[2], Jean-Michel Neumann[1], Jeremy C. Smith[1], Alain Sanson[1,]**

[1]Département de Biologie Cellulaire et Moléculaire, Section de Biophysique des Protéines et des Membranes and URA CNRS 2096, CEA Saclay, F-91191 Gif sur Yvette Cedex, France
[2]Computer Aided Engineering Laboratory, Research Institute in Informatics, Bucharest 71316, Romania

Received: 10 May 1998 / Accepted: 4 August 1998 / Published online: 2 November 1998

Abstract. Annexin molecules consist of a symmetrical arrangement of four domains of identical folds but very different sequences. Nuclear magnetic resonance (NMR) experiments on the isolated domains of annexin I in aqueous solution have indicated that domain 1 retains its native structure whereas domain 2 unfolds. Therefore these two domains constitute interesting models for comparative simulations of structural stability using molecular dynamics. Here we present the preliminary results of molecular dynamics simulations of the isolated domain 1 in explicit water at 300 K, using two different simulation protocols. For the first, domain 1 was embedded in a 46 Å cubic box of water. A group-based non-bonded cut-off of 9 Å with a 5–9 Å non-bonded switching function was used and a 2 fs integration step. Bonds containing hydrogens were constrained with the SHAKE algorithm. These conditions led to unfolding of the domain within 400 ps at 300 K. In the second protocol, the domain was embedded in a 62 Å cubic box of water. An atom-based non-bonded cut-off of 8–12 Å using a force switching function for electrostatics and a shifting function for van der Waals interactions were used with a 1 fs integration step. This second protocol led to a native-like conformation of the domain in accord with the NMR data which was stable over the whole trajectory (~2 ns). A small, but well-defined relaxation of the structure, from that observed for the same domain in the entire protein, was observed. This structural relaxation is described and methodological aspects are discussed.

Key words: Protein denaturation – Annexins – Domain stability – Molecular dynamics simulation

1 Introduction

For a complete understanding of the relationship between protein sequence and structure it is necessary to understand the mechanisms by which proteins fold and unfold. Calculations and experiments on small proteins in vitro can provide a lot information on these processes.

The annexins are a family of proteins which are of particular interest in folding studies. They are found in various insect, mammalian and plant tissues [1]. Although their biological function is only poorly understood, there is some evidence that they may be involved in membrane fusion, exocytosis, inflammation and blood coagulation [2]. The first stage of these phenomena involves membrane binding of the annexins in a calcium-dependent manner. The crystal structures of annexin I [3] and its domain 1 are depicted in Fig. 1. The interest of annexins for folding studies stems from the fact that X-ray structural analysis has revealed a hierarchical structural organization, involving one type of secondary structure, the alpha helix, one basic supersecondary pattern the helix-loop-helix or helix hair-pin, one specific tertiary fold, the domain : helix A-loop-helix B-helix C -helix D-helix E, and one type of circular association of domains into two modules, D2-D3 and D1-D4 [4, 5]. In the modules, the domains interact through the hydrophobic interface made by their helices B and E. Domains 2 and 4 interact via a hydrophilic interface.

The hierarchical simplicity of the annexin structure renders the examination of its folding pathways of particular interest. Therefore we have initiated a programme of research examining the stability, structure and dynamics of annexin fragments in vitro.

Within an annexin molecule, the domains have generally only about 30% of sequence homology but are structurally virtually identical. It was therefore particularly interesting to find that the isolated domains do not have identical folding properties [6–8]. In particular for annexin I, whereas the isolated domain 1 remains

*Contribution to the Proceedings of Computational Chemistry and the Living World, April 20–24, 1998, Chambery, France

**Also from Université P. et M. Curie, 9 Quai Saint-Bernard, Bât. C, 75005 Paris, France

Correspondence to: A. Sanson
Tel.: +33 1 69 08 28 63, Fax: +33-1-69 08 81 39,
e-mail: sanson@dsvidf.cea.fr

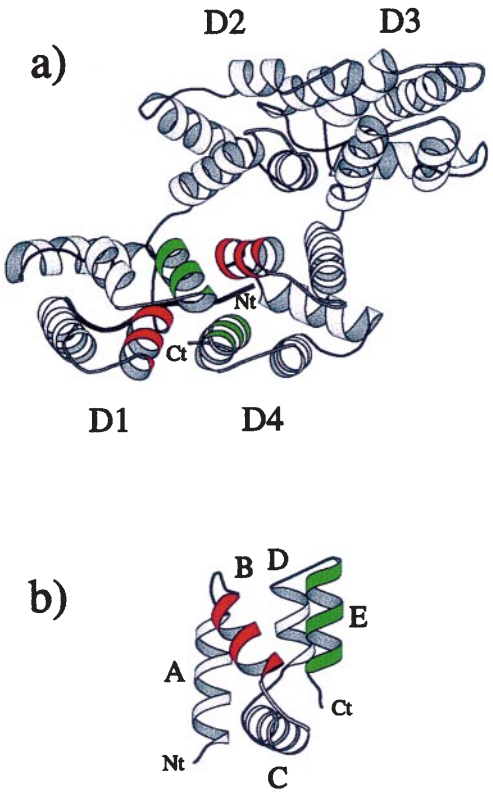

Fig. 1. **a** X-ray crystal structure of annexin I showing hierarchical structural organization of annexin domains and **b** structure of the domain 1; the names of the five helices are indicated. The B- and E-helices, which form the hydrophobic interdomain interface, are indicated in *red* and *green* respectively

structurally stable in aqueous solution the isolated domains 2 and 3, as well as the D2-D3 module do not. No information is presently available for domain 4 which remained inaccessible in the inclusion bodies of the bacteria from which the domains are extracted. This makes annexin domains interesting models for studying protein stability and unfolding with molecular simulation.

Molecular dynamics (MD) allow the intriguing possibility of being able to provide complete structural and dynamical information at the atomic level. This is of special interest when experimental data are missing or cannot be available. However the confidence that we have in these simulations must be assessed. The present work concentrates on this problem by investigating the effect of variation of the simulation protocol on the structural stability of domain 1. Domain 1 is a particularly good system with which to investigate simulation protocol for the following two reasons: (1) it is structurally similar to the native protein as described in [8], with a topology closely resembling that of the native X-ray structure (Fig. 1); (2) it is of only low stability relative to typical native protein of similar size, as evidenced from chemical denaturation studies [8]. Consequently, the structural integrity of the domain might be particularly sensitive to the simulation protocol, with small perturbations from ideal simulation conditions leading to unfolding of protein. It is this possibility that is tested in the present paper, by performing simulations of domain 1 with two different protocols. In one set, domain 1 unfolds significantly at 300 K in a few hundred picoseconds. In a second set of conditions the domain remains stable after a short period of reorganization that will be described.

The second aspect of the present work involves the examination of structural rearrangement accompanying interface removal. Applying MD simulation techniques to isolated annexin domains may also be useful for understanding structural and energetic aspects of domain-domain interactions in protein stability. Creation of a hydrophobic interface by dimerization is a known mechanism by which proteins increase their stability [9]. Because of the structural identity between annexin domains, formation of interdomain modules is somewhat analogous to dimerization: a hydrophobic interface is made by the two symmetrical pairs of B- and E-helices between domains 1 and 4 forming the D1-D4 module (Fig. 1). In this respect D1 is all the more interesting in that D1-D4 is a non-covalent module. With the aim of obtaining further insight into the domain interactions, we use the MD to examine possible structural relaxation of the domain accompanying excision from the entire protein.

2 Methods

2.1 Starting structure

The heavy atom coordinates of the starting structures were obtained by X-ray crystallography [3] (Kim, personal communication). Domain 1 was considered to consist of residues Thr 41 to Lys 113 (residues 14 to 86 in the annexin V numbering). The N-terminal segment of the protein (residues 1–42) was not included because it is topologically unrelated to domain 1 and, in addition, does not contribute to the domain stability (unpublished data). All protein atoms were explicitly included in the simulations. The hydrogen atoms were added using the HBUILD routine of CHARMM [10]. The C-ter and N-ter residues were not capped and remained charged as in the experimentally examined domain. The resulting protein was thus neutral. Both MD simulations were performed in the microcanonical ensemble with explicit solvent and periodic boundary conditions.

2.2 Simulation Method 1

A first MD simulation of the domain 1 was performed for 1100 ps using the CHARMM program version 23 with an all-atom force field (parm22). Domain 1 was embedded in a 46 Å cubic box of TIP3P water. Water molecules overlapping the protein were removed. The resulting domain 1 system contained 10 061 atoms: 1162 protein atoms, 2966 water molecules and one chloride counterion. Bonds containing hydrogens were constrained with the SHAKE algorithm [11]. The non-bonded interactions were calculated using a group-based cutoff and were smoothly brought to zero by multiplying by a cubic switching function between 5 and 9 Å. The relative dielectric constant ε was 1 and the integration time step was 2 fs. The simulation was carried out with the following protocol. The protein domains were fixed and the water molecules subjected to 1000 steps of steepest descent energy minimization and 1000 steps of adopted basis Newton-Raphson minimization. The water was then heated and equilibrated for 10 ps.

Subsequently, three further minimizations were carried out: first on the solvent molecules with the protein fixed, then on the protein with the solvent fixed, then on the whole system. The energy-minimized system was heated and equilibrated. Mass-weighted harmonic positional constraints were applied to the protein atoms during minimization, heating and equilibration. Heating to 300 K was over 30 ps. The force constants were 5.0 kcal/mol/Å^2 during the minimization, heating and the first 10 ps of equilibration. During the subsequent 45 ps of equilibration the force constants were gradually reduced to zero. At the end of the equilibration the Root Mean Square (RMS) deviation of the backbone heavy atoms from the crystal structure was 1.56 Å. The subsequent production phase was performed in the microcanonical ensemble without constraints for 1100 ps. Coordinates and velocities were saved from each simulation every 0.2 ps. The temperature was 300 K ± 10 K. Total computer (CPU) time for the simulation was about 1 month on an IBM 39H workstation.

2.3 Simulation Method 2

Two simulations were made with this method: the first starting from the X-ray structure involving a production of 1.2 ns and the second, starting from the minimized structure obtained from the 1.2 ns point of the first simulation, followed by further involving production of 500 ps. Both simulations using method 2 were atom-based with a non-bonded cutoff of 8–12 Å using a force switching function for electrostatics and a potential shifting function for van der Waals interactions, a cutoff for the non-bonded list of 14 Å and $\varepsilon = 1$ for electrostatics [12]. The integration time step was 1 fs without using the SHAKE algorithm. The CHARMM program, version 25b1, with the all-atom force field parameter set parm22, was run on a Cray T3E parallel computer. CPU time for the 1600 ps simulation was about 14 days with 64 processors.

The domains was embedded in a 62 Å cubic box of water. Water molecules with oxygen atoms within 2.8 Å of any non-hydrogen atoms of the protein were removed. The resulting system contained 23 831 atoms: 1166 protein atoms and 7555 water molecules. The following minimization procedure was used: (1) the protein was fixed while the solvent was minimized with 8000 steps of steepest descent, then the solvent was heated to 50 K (5000 steps), thermalized by 5×1000 steps of dynamics at 50 K, cooled to 20 K and finally minimized with 4000 steps of steepest descent; (2) the solvent was fixed, the constraints on the protein were removed and the protein was subjected to 2000 steps of steepest descent minimization; (3) all constraints were removed and the full system was minimized by 4000 steps of steepest descent.

2.3.1 First simulation

Heating to 300 K was performed at a rate of 10 K/ps. Equilibration was performed in two stages: in the first stage of 50 ps velocities were assigned using a Gaussian distribution and in the second stage the velocities were scaled every 50 steps to keep the temperature at 300 K ± 10 K. At the end of the equilibration the RMS deviation of the backbone heavy atoms from the crystal structure was 0.35 Å. During the 1600 ps production the temperature was 300 K ± 3 K and the total energy remained constant to within ± 0.05% fluctuation.

2.3.2 Second simulation

The system after the 1.2 ns of the previous 300 K simulation was annealed and minimized. This minimized structure was then used as a new starting structure and heated to 300 K and equilibrated as described above. At the end of the equilibration, the RMS deviation of the Cα atoms from the new starting structure was only 0.1 Å. During the subsequent 500 ps production the temperature was 300 K ± 3 K and total energy remained constant to within ± 0.05% fluctuation.

3 Results and discussion

3.1 Simulation with method 1: partial unfolding of the domain

The domain remained compact during this simulation, the radius of gyration increasing slightly from 11.5 Å to about 13 Å. The native overall chain topology, i.e. the relative positions in space of the chain segments that form helices in the native state, remained intact during the simulation. Nevertheless the domain substantially unfolded, losing a major part of its secondary structure after 1 ns of simulation.

In Fig. 2, curve (a) presents the Cα RMS deviation of domain 1 from the starting structure as a function of time. After the equilibration stage, the RMS deviation of domain 1 is already 1.6 Å, and exhibits a rapid linear increase up to ~4.5 during the first 400 ps followed by a more gradual increase to 5.5 Å at 1100 ps. The structure of the domain at 400 ps and 1000 ps is shown in Fig. 3. In the 400 ps structure a rather dense hydrophobic core, including the residues of the original core, remains in place. However, little of the original secondary structure remains consisting of a few clearly identifiable helix turns. In the 1000 ps structure the hydrophobic core is reduced in size and the secondary structure has almost completely disappeared.

The unfolding of the domain at 300 K, and in particularly the loss of the secondary structure is likely to be an artefact for several reasons: (1) as already mentioned, experimentally the isolated domain remains folded at 300 K and (2) NMR analysis of domain 1 fragments demonstrates that most of sequences encompassing the helix segments have a relatively high helix propensity (unpublished data). Protein fragments, devoid of the long-range interactions present in the native structure, are good reporters of the intrinsic structural propensities of sequences. The strong helix unfolding during the simulation is therefore likely to be erroneous.

3.2 Simulations with method 2: relaxation of the structure into a stable folded state

The second simulation protocol, described in Sect. 2, was that determined by comparing several simulations on a shorter peptide, the A-helix of domain 2 (Huynh et al., work in progress), with a large quantity of structural information from NMR [13]. This helix has a particular structural feature which was found to be very sensitive to the simulation protocol and especially to the way in which electrostatic interactions are treated. Among the methods tried, including Ewald summation, only the method described here (and used in Ref. [10]) was able to reproduce the experimental data. Full details of these methodological tests will be given in a later publication.

3.2.1 First simulation

The Cα RMS deviation of domain 1 from the starting structure as a function of time is presented in Fig. 2, curve (b). After the equilibration stage, the RMS

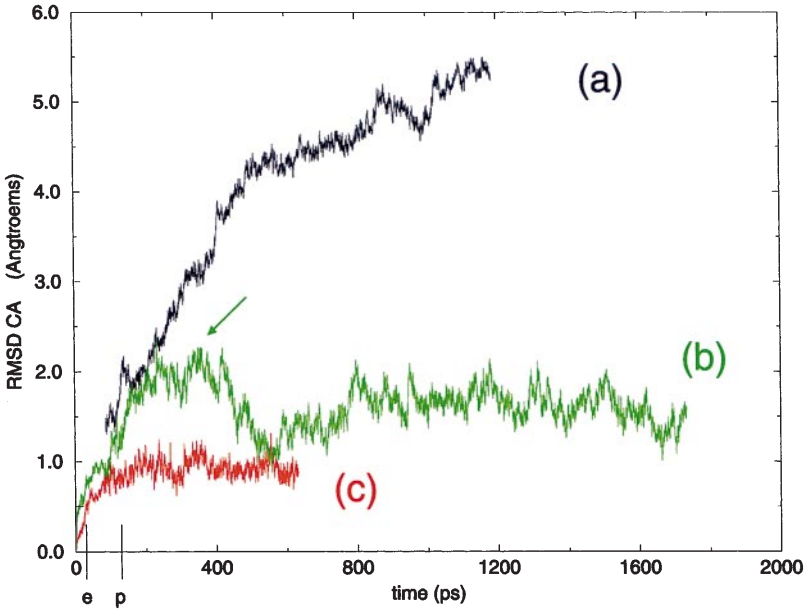

Fig. 2a–c. Time-dependence of the RMS deviation from the starting structure of the Cα atoms of domain 1: **a** simulation with method 1; **b** first simulation with method 2; **c** second simulation with method 2. The *arrow* indicates the point of maximum displacement of helix E and of maximum total solvent accessibility. Starting of equilibration and production are indicated with *e* and *p*, respectively

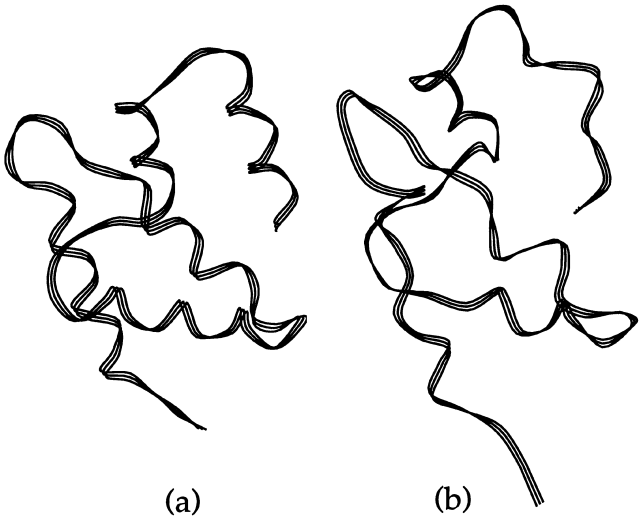

Fig. 3a, b. Ribbon representation of domain 1 after 400 ps of the simulation performed using method 1: **b**. The starting structure **a** for this simulation was the crystallographic structure of the domain in the entire protein

deviation of domain 1 remains small at 0.35 Å. During the first 400 ps of production, the RMS deviation increases up to a maximum of about 2 Å after ~250 ps then decreases and stabilizes to a constant value of about 1.8 Å, a value in the range commonly seen for simulation of native proteins at 300 K. The 2 Å maximum, which also corresponds to a point of maximum solvent accessibility, accompanies a structural relaxation of the domain after excision from the entire protein. This relaxation will be described later on.

3.2.2 Second simulation

The structure after 1.2 ns was annealed and minimized (see methods) in order to create a new starting structure, i.e. that of the isolated domain, to be compared to the first starting structure, i.e. the crystallographic structure of the domain in the entire protein. The RMS deviation of the Cα atoms from the first starting structure is ~2.3 Å. The new starting structure was also heated to 300 K and equilibriated. At the end of the equilibriation stage, the RMS deviation of the Cα atoms from the new starting structure was only 0.1 Å, and rapidly reached a rather small plateau value of about 0.9 Å as shown in Fig. 2, curve (c). The relative structural stability observed during this second simulation suggests that this structure represents an equilibrium conformation of the isolated domain 1 in aqueous solution.

3.3 Description of the structural relaxation

The structural relaxation may be essentially visualized as a rigid-body movement of helix E and, to a lesser extent of helix D, with respect to the A-, B- and C-helices. This is illustrated in Fig. 4. Helix E first moves, upwards in the figure, during the first 300 ps allowing the sliding of its hydrophobic residues on the remaining part of the hydrophobic core. The helix then moves laterally to adapt the contact with the hydrophobic core and finally moves slightly downwards. The result is a shift of about 3 Å in the staggered arrangement of the hydrophobic side chains of helix E against the rest of the hydrophobic core. Helix D also moves aside by approximately 1.5 Å. The native hydrophobic interface with domain 4 contains residues Ile26 and Thr30 of helix B and Leu64, Val67 and Leu71 of helix E. The net effect of the structural relaxation on these residues is that their solvent accessibility is reduced from ~175 $Å^2$ to ~125 $Å^2$ for helix B and from ~160 $Å^2$ to ~125 $Å^2$ for helix E. The change in solvent accessibility is clearly depicted in Fig. 5, which shows the effective displacement of the hydrophobic residues of the helix E and their partial

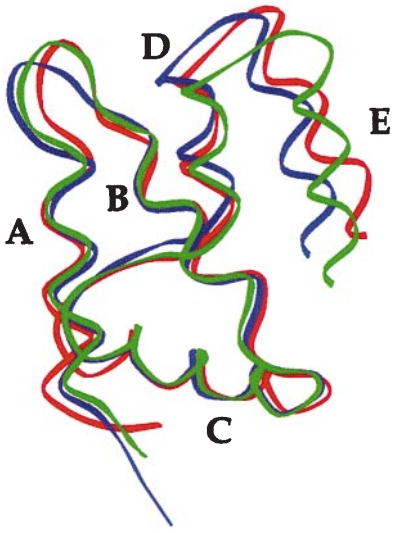

Fig. 4. Structural relaxation of domain 1 after excision from the native protein, observed using method 2: starting structure (*green*); snapshot after ~250 ps of production (*arrow* in Fig. 2), the point of maximum displacement of the helix E (*red*); minimized final structure of the isolated domain (*purple*)

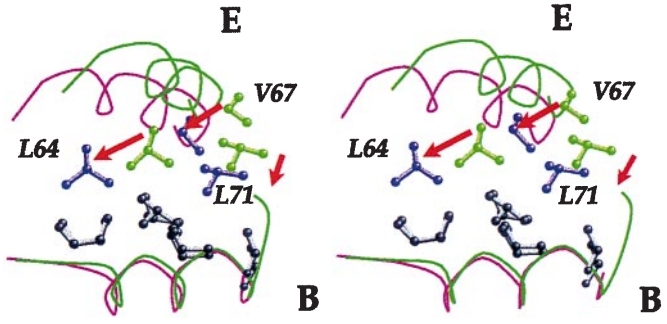

Fig. 5. Close up view of helices B and E of the domain in the crystallographic starting structure (*green*), and in the final structure of the isolated domain obtained with method 2 (*purple*). Helix B in the two structures is superposed. The black residues are three core residues, stable in space, in the two structures and the red *arrows* indicate displacements of residues L64, V67 and L71

burial. The figure also shows the new arrangement of helix E, more parallel to helix B.

4 Conclusion

The work presented here demonstrates the structural sensitivity of the annexin I domain 1 to the MD simulation conditions. The simulations were performed on the 10^{-10}–10^{-9} s time scale in explicit water. As such these calculations are computationally expensive. It is therefore not practical to painstakingly systematically vary elements of the simulation methodology one by one so as to identify which are responsible for any given effect. Consequently there are several differences between method 1 and method 2. These include the method of smoothing the long-range electrostatic interactions, the simulation-box size and the integration time step. What is clear however is that the two methods have produced markedly different results on the structural stability of the isolated annexin domain. This was not entirely expected as simulation method 1, which led to an experimentally unobserved rapid unfolding of the domain, has previously been used with success in the simulation of protein structure and dynamics. One reason for the present unfolding may well be the stability of the folded isolated domain, which experimentally is low compared to many small native proteins. This may render the protein structure particularly sensitive to errors.

Method 2 clearly produces a protein structure stable over the nanosecond time scale that is close to the experimentally determined structure. Consequently, the chance is increased that the structural change observed may be real. Of particular interest in the present work is the structural rearrangement at the domain interface seen on excision of the domain. This involves shifting of helices and associated burial of hydrophobic groups. Confirmation of this structural change awaits further high-resolution NMR work.

References

1. Moss SE (1992) The Annexins. Portland, London
2. Raynal P, Pollard HB (1994) Biochim Biophys Acta 1197:63
3. Weng X, Luecke H, Song IS, Kang DS, Kim SH, Huber R (1993) Protein Sci 2:448
4. Huber R, Berendes R, Burger A, Schneider M, Karshikov A, Luecke H, Romisch J, Paques E (1992) J Mol Biol 223:683
5. Favier-Perron B, Lewit-Bentley A, Russo-Marie F (1996) Biochemistry 35:1740
6. Cordier-Ochsenbein F, Guerois R, Baleux F, Huynh-Dinh T, Chafotte A, Neumann JM, Sanson A (1996) Biochemistry 35:10347
7. Cordier-Ochsenbein F, Guerois R, Baleux F, Co YM, Huynh-Dinh T, Lirsac PN, Russo-Marie F, Neumann JM, Sanson A (1998) J Mol Biol 279:1163–1175
8. Cordier-Ochsenbein F, Guerois R, Russo-Marie F, Neumann JM, Sanson A (1998) J Mol Biol 279:1177–1185
9. Jaenicke R (1991) Eur J Biochem 202:715
10. Brooks BR, Bruccoleri M, Olafson BD, States DJ, Swaminathan S, Karplus M (1983) J Comp Chem 4:187
11. Ryckaert JP, Cittotti G, Berendsen HJC (1977) J Comp Phys 23:327
12. Steinbach PJ, Brooks BR (1994) J Comp Chem 15:667
13. Guérois R, Cordier-Ochsenbein F, Baleux F, Huynh-Dinh T, Neumann JM, Sanson A (1998) Protein Science 7:1506–1515