doi:10.1016/j.jmb.2004.04.007 J. Mol. Biol. (2004) 339, 515–525

JMB

Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER

Investigating the Accessibility of the Closed Domain Conformation of Citrate Synthase using Essential Dynamics Sampling

Isabella Daidone[1], Danilo Roccatano[2]* and Steven Hayward[3]*

[1]*Department of Chemistry University of Rome, "La Sapienza" P.le A. Moro 5 00185 Rome, Italy*

[2]*School of Engineering and Science, International University Bremen, Campus Ring 1, D-28725 Bremen Germany*

[3]*School of Computing Sciences and School of Biological Sciences, University of East Anglia, Norwich NR4 7TJ, UK*

**Corresponding authors*

A molecular dynamics study of pig heart citrate synthase is presented that aims to directly address the question of whether, for this enzyme, the ligand-induced closed domain conformation is accessible to the open unliganded enzyme. The approach utilises the technique of essential dynamics sampling, which is used in two modes. In exploring mode, the enzyme is encouraged to explore domain conformations it might not normally sample in free molecular dynamics simulation. In targeting mode, the enzyme is encouraged to adopt the domain conformation of a target structure. Using both modes extensively, it has been found that when the enzyme is prepared from a crystallographic open-domain structure and is in the unliganded state, it is unable to adopt the crystallographic closed-domain conformation of the liganded enzyme. Likewise, when the enzyme is prepared from the crystallographic closed liganded conformation with the ligands removed, it is unable to adopt the crystallographic open domain conformation. Structural investigations point to a common structural difference that is the source of this energy barrier; namely, the shift of α-helix 328–341 along its own axis relative to the large domain. Without this shift, the domains are unable to close or open fully. The charged substrate, oxaloacetate, binds near the base of this helix in the large domain and the interaction of Arg329 at the base of the helix with oxaloacetate is one that is consistent with the shift of this helix in going from the crystallographic open to closed structure. Therefore, the results suggest that without the substrate the enzyme remains in a partially open conformation ready to receive the substrate. In this way, the efficiency of the enzyme should be increased over one that is closed part of the time, with its binding site inaccessible to the substrate.

Keywords: domain motion; molecular dynamics; principal component analysis; DynDom analysis

Introduction

Domain movements form a large class of functional movements in enzymes, and some effort has been made to understand and characterise them.[1–3] In the simplest scenario, the substrate binds to an open conformation of the enzyme inducing closure. Once closed, the reaction catalysed by the enzyme can proceed in a protected and highly specific environment. It is of interest to know whether the closed domain conformation of the enzyme is accessible or inaccessible to the unliganded open conformation of the enzyme†. Gerstein *et al*.[2] in their review of domain movements speculate that both the open and closed conformations are dynamically accessible to the unliganded enzyme at physiological temperature. Their model implies that there exists a continuous range of stable domain conformations between the most open and closed ones. They base their

Abbreviations used: MD, molecular dynamics; EDS, essential dynamics sampling; RG_RMSD, rigid body root mean-square deviation.

E-mail addresses of the corresponding authors: d.roccatano@iu-bremen.de; sjh@cmp.uea.ac.uk

† We use the word inaccessible to indicate a free energy barrier between states, making it very unlikely for there to be a transition from one to the other.

arguments partly on the finding that a closed unliganded form of the binding protein lactoferrin is stabilised by weak crystal packing forces.[4] Although this may be true for some domain proteins, it is not necessarily a universal truth, as many domain proteins have much more complicated domain interfaces than lactoferrin. Indeed, recent combined NMR and fluorescence experiments on maltose-binding protein have confirmed a barrier between the open and closed domain conformations for that protein.[5] Crystallographic work on citrate synthase does not support the idea of a continuous range of stable domain conformations but one where there are just two stable states related to enzymatic mechanism.[6] Molecular dynamics (MD) simulations on pig heart citrate synthase suggest that there is a large free energy barrier to surmount to reach the crystallographic closed domain conformation from the crystallographic open conformation in the unliganded state.[7] On the basis of that study, it was concluded that the energy to surmount this barrier comes from the interaction of the enzyme with the substrate. The concept of an energy barrier between the open and closed domain conformation would make sense for an enzyme, in that an enzyme that remained open would be more efficient than one that spent some of its time closed with its binding site inaccessible to the substrate.

Citrate synthase catalyses a step in the citric acid cycle, namely the Claisen condensation of acetyl-coenzyme A with oxaloacetate to form citrate and coenzyme A.[8] It is a homodimer, where the monomer comprises a large and a small domain. It is an enzyme that displays a classic domain movement as part of its function, where the binding of oxaloacetate induces domain closure, upon which the binding site for acetyl-coenzyme A is formed.[8] The MD simulation study referred to above comprised three simulations that started from the crystallographic open conformation and a further three simulations that started from the crystallographic closed conformation. In both cases, any ligands were removed. The simulations starting from the open conformation appeared to show that first, there are a large number of unliganded domain conformations that are accessible from the unliganded open conformation, but second, the crystallographic closed domain conformation cannot be reached from the crystallographic open conformation in the unliganded state. In the simulations starting from the crystallographic closed conformation, the trajectories remained around the closed domain conformation, apart from one, where one of the monomers made an apparently spontaneous transition to the region explored by the open simulations. Although these results suggested that the closed conformation could spontaneously convert to the open in the absence of the products, this probably never occurs in reality because the enzyme needs first to open to allow citrate to escape. In order to investigate the results of the previous free MD simulations further, in this work we have used the essential dynamics sampling (EDS) technique.[9] This technique has been used in a number of studies on protein[10] and peptides dynamics[11] as well as folding/unfolding simulations,[12,13] but this is its first application to the study of a functional domain movement in an enzyme. Using this technique, we are able to encourage the domain conformation to explore a larger region of space than it would in free MD, and to move towards particular target conformations.

Results

Simulations from open conformation

Exploring mode simulations

Three exploring mode simulations were performed starting from the open conformation that was used to start production runs in the original work.[7] This open domain conformation does not coincide exactly with the crystallographic open domain conformation but is very near to it in comparison to the crystallographic closed domain conformation. Details of these simulations are given in Table 1. Figure 1 shows the domain trajectories projected onto the two main degrees of freedom for the domain movement (see Methods). Also shown in this Figure are the trajectories of the original free simulations,[7] starting from the open and closed. It is clear that domain conformations are explored that are not explored in the original

Table 1. Runs originating from crystallographic open structure

Run	Mode	Starting structure	Length of simulation (ps)
Run 1	Exploring	Equilibrated crystal open[a]	1300
Run 2	Exploring	Equilibrated crystal open[a]	1165
Run 3	Exploring	Equilibrated crystal open[a]	500
Run 4	Target to closed	Equilibrated crystal open[a]	500
Run 5	Target to closed	Equilibrated crystal open[a]	500
Run 6	Target to closed	Equilibrated crystal open[a]	500
Run 7	Target to closed	Conformation at 800 ps of run 1	500
Run 8	Exploring	Final conformation of run 5	500
Run 9	Target back-to-open	Conformation at 40 ps of run 8	500

[a] Each had a unique set of starting velocities generated from the Maxwell distribution.

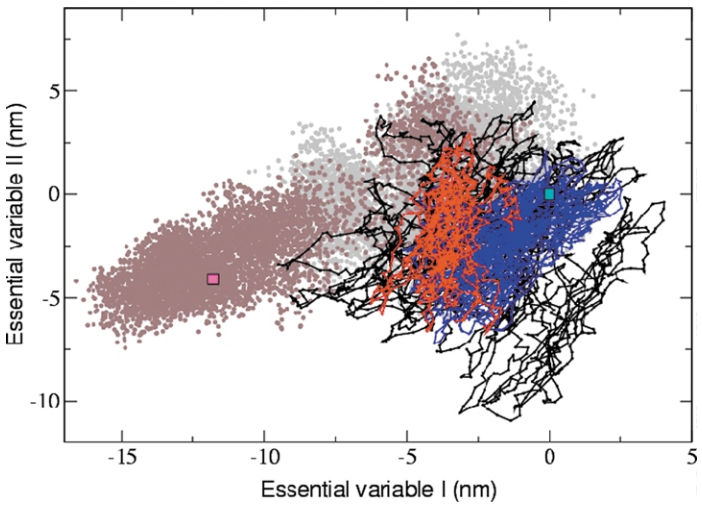

Figure 1. Projections of the trajectories of the exploring mode simulations that started from the crystallographic open conformation, onto the first two eigenvectors of the rigid-body essential dynamics analysis. The three simulations shown in black, blue and red correspond to run 1, run 2 and run 3 of Table 1, respectively. The crystallographic open and closed conformations are indicated with a cyan and magenta filled square, respectively. The trajectories of the original free simulations[7] starting from the crystallographic open and closed conformations are shown in grey and brown, respectively.

open free simulations. In run 1 there is some probing towards the crystallographic closed domain conformation, but generally the region around it is avoided. In order to investigate further whether trajectories that start from the crystallographic open conformation are able or unable to reach the crystallographic closed domain conformation, targeted simulations were performed.

To-closed targeting mode simulations

In these four simulations, the crystallographic closed domain conformation is the target. Three simulations started from the same open conformation from which the exploring mode simulations were started, although each had a different set of velocities derived from the Maxwell velocity distribution. The fourth simulation started from the 800 ps conformation of run 1 of the exploring mode simulations. Table 1 gives the details of these runs and Figure 2 shows their domain trajectories. All trajectories move rapidly towards the target but eventually are unable to move any closer and remain stuck around a region of closest approach. The simulation with the different starting conformation gets stuck in a slightly different region from the others. These simulations confirm that the crystallographic closed domain conformation is inaccessible to conformations that start from the crystallographic open conformation. This finding supports the conclusion from the earlier work,[7] where a free energy barrier between the open and closed was proposed.

Further simulations

Figure 2 shows an exploring simulation that was started from the final conformation of a targeting simulation. In addition, a simulation that targeted

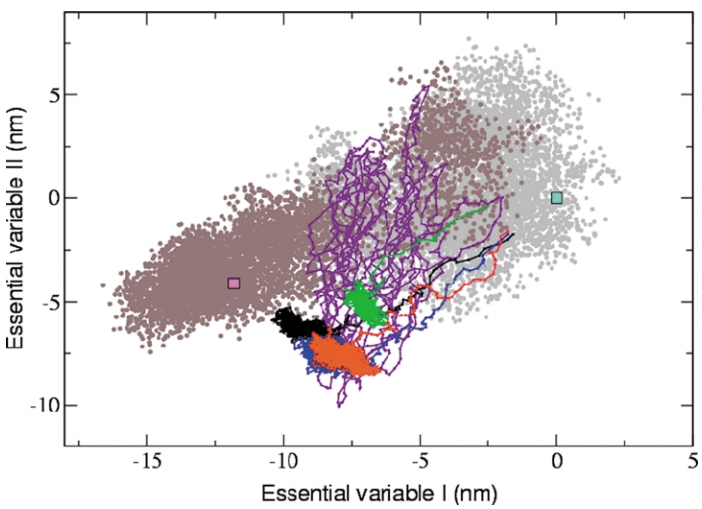

Figure 2. Projections of the trajectories of the "to-closed" targeting mode simulations starting from the crystallographic open conformation, onto the first two eigenvectors of the rigid-body essential dynamics analysis. The four simulations shown in black, blue, red and green correspond to run 4, run 5, run 6 and run 7 of Table 1, respectively. The projection of the exploring mode simulation corresponding to run 8 of Table 1 is shown in violet. The crystallographic open and closed conformations are indicated with a cyan and a magenta filled square, respectively. The trajectories of the original free simulations[7] starting from the crystallographic open and closed conformations are shown in grey and brown, respectively.

Table 2. Runs originating from crystallographic closed structure

Run	Mode	Starting structure	Length of simulation (ps)
Run 1	Exploring	Equilibrated crystal closed[a]	500
Run 2	Exploring	Equilibrated crystal closed[a]	500
Run 3	Exploring	Equilibrated crystal closed[a]	500
Run 4	Target to open	Equilibrated crystal closed[a]	500
Run 5	Target to open	Equilibrated crystal closed[a]	500
Run 6	Target to open	Equilibrated crystal closed[a]	500
Run 7	Target to open	Conformation at 500 ps of run 2	500
Run 8	Exploring	Final conformation of run 4	850
Run 9	Target back-to-closed	Conformation at 500 ps of run 8	500

[a] Each had a unique set of starting velocities generated from the Maxwell distribution.

back to the crystallographic open domain conformation was performed. The trajectory did not reach the crystallographic open domain conformation. Details of these simulations are given in Table 1.

Simulations from closed conformation

Exploring mode simulations

Three exploring mode simulations were performed from the closed starting conformation, which was the same starting point as for the free MD simulations and was close to, but not coincident with, the crystallographic closed domain conformation. Details of the simulations are reported in Table 2. The projected domain trajectories are displayed in Figure 3. They move in a region around the main distribution of closed domain conformations determined in the free simulations. However, as one would expect in an exploring mode simulation, they went beyond some of the outermost regions explored in the original closed simulations.

To-open targeting mode simulations

In these four simulations, the crystallographic open domain conformation is the target. Three simulations started from the same closed conformation, although each had a different set of velocities derived from the Maxwell velocity distribution. The fourth simulation started from the 500 ps conformation of run 2, an exploring mode simulation. Table 2 gives the details of these simulations and Figure 4 shows their domain trajectories. All the trajectories move rapidly towards the target but eventually are unable to move any closer and remain stuck around a region of closest approach. These simulations appear to show that the crystallographic open domain conformation is inaccessible to conformations that start from the closed. These trajectories follow the path taken by the "transitional trajectory" of the free MD simulation study,[7] indicating the consistency of these results with that study. However, from the free MD study, it was concluded that the transitional trajectory was able to reach the open conformation. This conclusion is not supported by the results presented here. The domain conformations at the end of these targeted trajectories are indeed closer to the crystallographic open domain conformation than the closed, but some internal differences prevent these conformations from reaching the crystallographic open domain conformation.

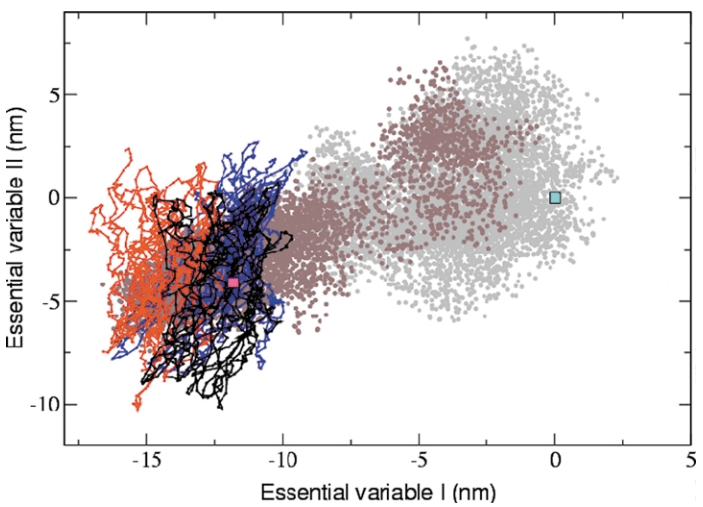

Figure 3. Projections of the trajectories of the exploring mode simulations starting from the crystallographic closed conformation, onto the first two eigenvectors of the rigid-body essential dynamics analysis. The three simulations shown in black, blue and red correspond to run 1, run 2 and run 3 of Table 2, respectively. The crystallographic open and closed conformations are indicated with a cyan and a magenta filled square, respectively. The trajectories of the original free simulations[7] starting from the crystallographic open and closed conformations are shown in grey and brown, respectively.

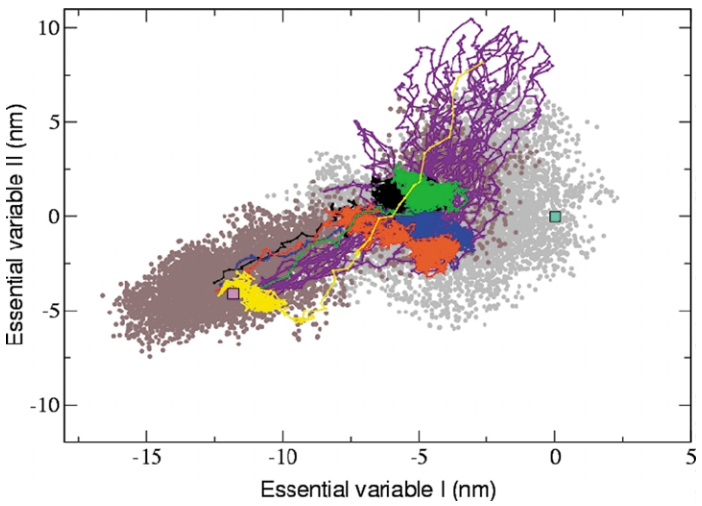

Figure 4. Projections of the trajectories of the "to-open" targeting mode simulations starting from the crystallographic closed conformation, onto the first two eigenvectors of the rigid-body essential dynamics analysis. The four simulations shown in black, blue, red and green correspond to run 4, run 5, run 6 and run 7 of Table 2, respectively. The projections of run 8 (exploring mode simulation) and run 9 (targeting mode simulation) of Table 2 are shown in violet and in yellow, respectively. The crystallographic open and closed conformations are indicated with a cyan and a magenta filled square, respectively. The trajectories of the original free simulations[7] starting from the crystallographic open and closed conformations are shown in grey and brown, respectively.

Further simulations

Figure 4 shows an exploring mode simulation (run 8 in Table 2) that was started from the end of the targeted simulation, run 4. Again, this trajectory explores the regions accessed by the transitional trajectory going back to the crystallographic closed domain conformation six times along the same path as the transitional trajectory. This supports the finding that the crystallographic open domain conformation is inaccessible and that there is an internal difference that allows these rather open domain conformations to reach the crystallographic closed domain conformation, but not the open. At the 500 ps conformation of this exploring mode trajectory, a targeted simulation (run 9, Table 2) was started with the crystallographic closed conformation as the target. The trajectory shown in Figure 4 confirms that the crystallographic closed domain conformation is indeed accessible from conformations originating from the crystallographic closed conformation. This was not found in the back-to-open simulation. The path taken in this back-to-closed trajectory is again the same as the transitional trajectory and the exploring mode trajectory of run 8, indicating a low-energy pathway.

Identifying the source of the energy barrier between open and closed conformations

These results confirm one finding of the previous work:[7] that the crystallographic closed domain conformation is inaccessible from the open unliganded conformation. However, on the basis of the transitional trajectory, it was speculated that the crystallographic open domain conformation is accessible from the crystallographic closed conformation in the unliganded state. However, this work indicates that the crystallographic open domain conformation is not accessible from the crystallographic closed conformation. This suggests that an energy barrier exists between both experimental domain conformations in their unliganded states. A barrier to domain rotation is surely located at an interface region situated between the two domains. This would mean that it is likely to be assigned as a bending region in the DynDom analysis. Given that this barrier is obviously overcome in the presence of the substrate or product, structural changes in the vicinity of the binding sites for these ligands are of particular interest. Recently, Hayward has developed a method to identify residues that are involved in inducing closure in enzymes when an open unliganded structure and a closed liganded structure are available.[14] Using this method, three potential "closure-inducing residues" have been identified in citrate synthase: His274, His320 and Arg329. These residues interact with the substrate oxaloacetate to induce closure and it is reasonable to expect, therefore, that the barrier to closure will be located in the vicinity of these residues. Previously, it was speculated that the barrier between the open and closed conformations is located at the ψ-dihedral of His274,[7] which in the closed conformation has an angle of $-134.7°$, which together with a ϕ-dihedral angle of $-114.7°$ puts it in a "disallowed" region of the Ramachandran plot.[15] However, a structure at 140 ps of the back-to-closed targeting simulation (run 9 in Table 2), has a domain conformation that is almost identical with that of the crystallographic closed domain conformation, but the ψ-dihedral angle of His274 has a value of $-77.6°$, which with a ϕ-dihedral angle of $-61.3°$ puts it in a low-energy region of the Ramachandran plot. It seems, therefore, that the extreme value of the dihedral angle in the crystallographic closed conformation is due to the interaction of this residue with the substrate

oxaloacetate, and that it need not have this value in order for the enzyme to reach the closed domain conformation. This means that there is no particular hindrance to the domain rotation from this region. Unlike His274, His320 is not assigned as a bending region and is located in a region that moves as a rather rigid body in going from the crystallographic open to closed conformation. As it is not located in an interdomain region, it is difficult to see how any barrier to domain closure resides in the vicinity of His320. Arg329 is assigned as a bending residue and is situated at the N terminus of an α-helix that undergoes a significant shift upon domain rotation. Its role in domain closure will be elucidated below.

Structural analysis of targeting simulations

The domain conformations in the targeting simulations are straining to achieve their target conformations but are unable to do so. It would appear logical, therefore, to analyse the movements between the starting conformations, the endpoint conformations of these targeting simulations, and the target conformations themselves. Table 3 gives the actual domain rotation angle corresponding to these movements and the rigid-body root mean-square deviations (RG_RMSDs). The data in Table 3 verify that there is a significant difference between the fully open and closed conformations as determined by crystallography, and the most open and closed conformations achieved in the targeting simulations. The difference between the domain movements from starting to endpoint conformations and endpoint to target conformations can be appreciated easily from Figures 2 and 4. The starting to endpoint domain movements take routes that are rather parallel with the line that joins the crystallographic conformations directly. The routes taken from the open are on the opposite side of this line to those that start from the closed (see Figure 5). The endpoint to target domain movements would be more perpendicular to this line. The symmetry implied by Figure 5 indicates a common structural mechanism that prevents the

Table 3. Domain rotation angles and RG_RMSDs

Targeting to closed from open	Small domain rotation angle between start and endpoint (deg.)	RG_RMSD between small domain at start and endpoint (Å)	Small domain rotation angle between endpoint and target (deg.)	RG_RMSD between small domain at endpoint and target (Å)
Run 1	20.3	5.6	2.5	1.8
Run 2	15.5	5.4	5.2	2.6
Run 3	14.4	5.2	7.7	3.1
Run 4	14.0	4.2	8.4	3.1
Targeting to open from closed				
Run 1	15.1	4.1	7.1	2.9
Run 2	13.6	4.5	5.8	1.8
Run 3	12.1	4.3	8.5	2.1
Run 4	16.7	4.4	7.1	3.0

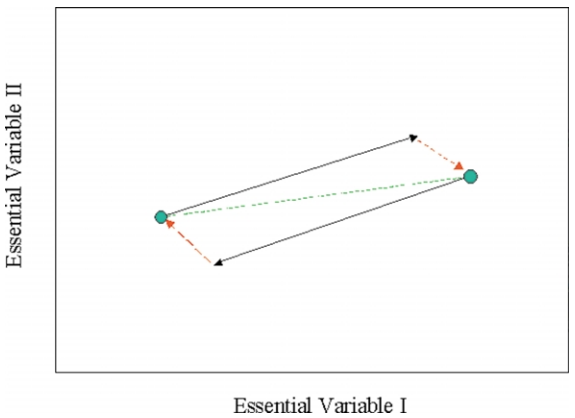

Figure 5. An illustration of the paths taken by the trajectories of the targeted simulations in relation to the locations of the crystallographic domain conformations (see Figures 2 and 4). The filled circles indicate the crystallographic domain conformations. The unbroken arrows indicate the general direction taken by the targeted trajectories, which start from the crystallographic domain conformation indicated at the arrow's origin. The broken arrows point to the targeted crystallographic domain conformation from the final conformation located at the head of the unbroken arrow. In both cases, the direction taken by the targeted simulations is rather parallel with, but does not follow the direct path between the two crystallographic domain conformations. In order to achieve these conformations, a movement indicated by the broken arrows is required. However, in both cases it appears that this movement is unable to occur.

conformations of the unliganded enzyme originating from the open crystallographic structure reaching the closed, and *vice versa*. In order to investigate this, two structures were selected that form a line on our two-dimensional (2D) projections that is parallel with the direction of the trajectories of the targeting simulations before they get stuck, and a further two that form a line that is perpendicular to this line. Then the program Dom_Select was used to characterise the rigid-body movement of the small domain relative to the large for both of these pairs. In Figure 6, these hinge axes are displayed with the enzyme structure. The hinge axis depicting the movement in the parallel direction lies directly between, and is almost perfectly parallel with, a pair of parallel α-helices: helix 222–235 situated in the large domain, and helix 328–341 situated in the small. The hinge axis depicting the perpendicular movement is not parallel with these helices, but is situated between them. The fact that in both cases the axes are located directly between these two helices indicates that they play a crucial role in the interdomain movement.

Shift of α-helix 328–341

Although helix 328–341 is assigned to the small domain here, DynDom often assigns a portion of this helix to belong to the large domain in terms

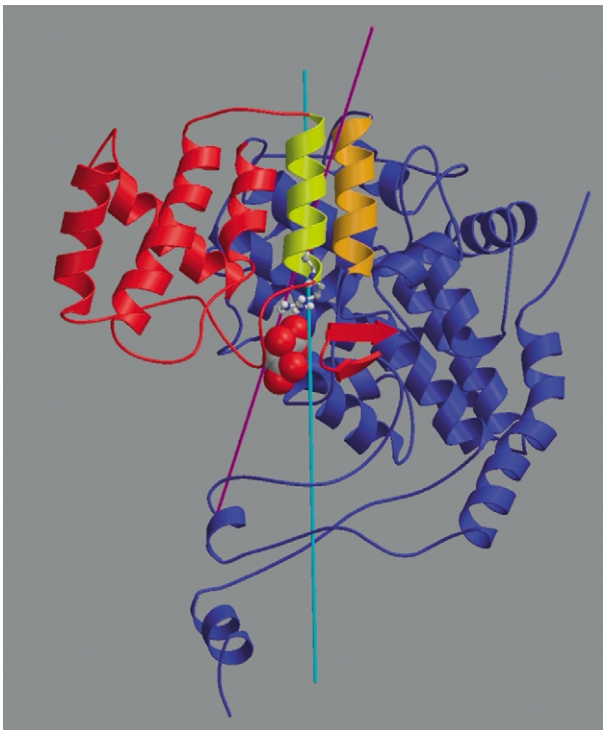

Figure 6. Backbone trace of the citrate synthase monomer. The large domain is coloured blue; the small domain is coloured red; α-helix 328–341 is coloured yellow; α-helix 222–235 is coloured orange. The cyan and magenta rods indicate the hinge axes for the domain movements indicated by the unbroken and broken arrows in Figure 5, respectively. These axes were calculated by selecting two pairs of conformations, one pair projecting parallel with the path taken in the targeted trajectories, and the other perpendicular to this path. These pairs of conformations were then passed to the program Dom_Select. The substrate, oxaloacetate, is fitted into the large domain and is depicted in space-filling model. Arg329, which is situated at the base of α-helix 328–341, is shown in ball-and-stick mode. In the crystallographic closed structure, Arg329 and oxaloacetate form a strong salt-bridge. It is thought that this interaction helps citrate synthase overcome the energy barrier in moving from the open to the fully closed state. The Figure was created using RasMol,[26] MOLSCRIPT[27] and Raster3D.[28]

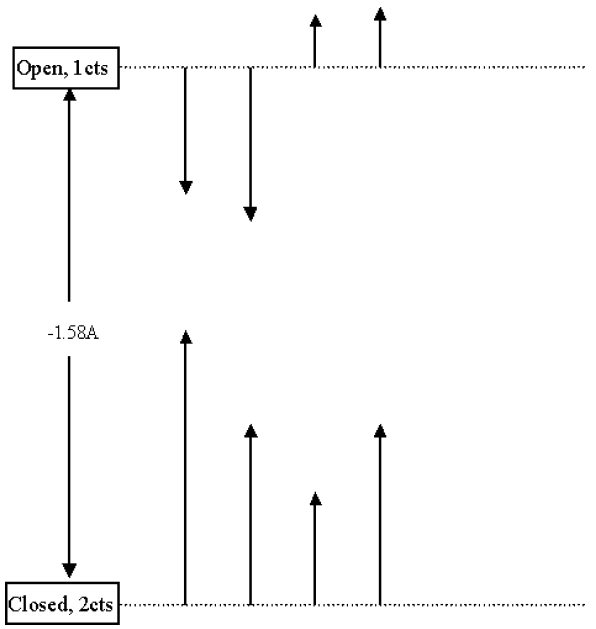

Figure 7. In going from the crystallographic open to closed structure the α-helix 328–341 shifts −1.58 Å along its own axis relative to the large domain. The arrows show the shifts of this helix relative to the large domain from the starting conformation in the targeted simulations. When starting from the open conformation and targeting to the crystallographic closed conformation, this helix is unable to shift down sufficiently. Likewise, when starting from the closed conformation and targeting to the crystallographic open conformation, this helix is unable to shift up sufficiently. The distances were calculated by the program Helix _Shift and the lengths of the arrows in the Figure correspond to the calculated distances.

of its rotational properties. (Please see the DynDom database of protein domain motions for more details on the domain movement of citrate synthase between crystallographic open and closed conformations.[16]) Helix 222–235 belongs unambiguously to the large domain. The movement between the crystallographic open and closed conformations is described by a hinge axis that makes an angle of approximately 30° with these helices. Consequently, the movement from crystallographic open to closed results in a distinct shift of the helix 328–341 "downwards" relative to the large domain (for convenience, the direction "up" will be used to refer to the direction of the helical axis of this helix pointing along the direction given by the right-hand rule). As the hinge axis describing the movement from starting to endpoint conformations in the targeting simulations is parallel with the helical axis, rotation about this hinge axis should not result in a shift of helix 328–341 relative to the large domain. The hinge axis describing the endpoint to target conformations, however, is not parallel with this helical axis, and rotation about this hinge axis would result in a shift of helix 328–341 relative to the large domain. Thus, the movement described by the crystallographic hinge axis comprises two movements (see Figure 5), one a rotation about an axis parallel with the two helices, which does not result in a relative shift of these helices, the other a rotation about an axis not parallel with the two helices, which does result in a relative shift of the helices. The former does occur in our simulations, but the latter cannot. In order to quantify the shift in helix 328–341 relative to the large domain, the program Helix_Shift was used (see Methods). In going from the crystallographic open to closed conformations, this helix shifts −1.58 Å along its own axis relative to the large domain. The shift of this helix in each of the targeting simulations was calculated as the shift in the helix between the starting and endpoint conformations. Figure 7 shows these

shift distances. It confirms that the helix is unable to move up or down sufficiently to reach the target conformation as the domains rotate. Thus, the shift of helix 328–341 relative to the large domain that occurs between the crystallographic open and closed conformations is unable to proceed sufficiently in our simulations. This result implies that the barrier to opening and closing is a barrier to the shift of this α-helix. Given that the open domain conformation does reach the closed domain conformation in the presence of the substrate oxaloacetate, it is probably the interaction with oxaloacetate that is able to shift helix 328–341 relative to the large domain. Arg329 is situated at the base of this helix and is often assigned as a bending residue. In a sequential model of ligand binding and domain closure, oxaloacetate binds first to the large domain before closure occurs.[14] This would then put oxaloacetate in a position to interact with Arg329 in the open conformation. This interaction creates a torque about the hinge axis helping to induce the closed conformation. In the closed conformation, the salt-bridge between Arg329 and oxaloacetate is fully formed. The suggested movement that this interaction would induce is one that would shift helix 328–341 downwards relative to the large domain (see Figure 6). Helix 328–341 is parallel with helix 222–235 in the large domain and they have many packing interactions. It would appear that the movement of helix 328–341 is quite constrained by these packing interactions and, therefore, the interaction between Arg329 and oxaloacetate is one that is consistent with these inter-helix contacts. Therefore, our hypothesis is that it is primarily the interaction between oxaloacetate and Arg329 that is able to shift this helix downwards over the energy barrier. However, there is no easy explanation as to how the presence of the product might shift this helix back to its fully up position in the open conformation as citrate forms a strong salt-bridge with Arg329 in the closed conformation.[15] If this hypothesis is correct, then the energy barrier resides in the interactions between the parallel helices 222–235 and 328–341.

Importance of Arg329

Arg329 is conserved over all available citrate synthase sequences despite some of the sequences having diverged considerably. This was ascertained by using the sequence retrieval system (SRS), at the European Bioinformatics' Institute† by selecting EC number 2.3.3.1 from the Enzyme database,[17] and then linking to the SWISS-PROT protein sequence database.[18] All the sequences found were then aligned using CLUSTAL W.[19] Arg329 is also the last residue in the citrate synthase PROSITE[20] motif:

G-[FYA]-[GA]-H-X-[IV]-X(1, 2)-[RKT]-X(2)-D-[PS]-R

In all available PDB structures of citrate synthase, this arginine residue is situated at the same structural position, namely at the N terminus of one of two parallel α-helices. In all closed structures, this arginine residue makes a salt-bridge with either citrate or oxaloacetate. Unfortunately, only two mutational studies of this residue are reported in the literature. Both these are on citrate synthase from *Escherichia coli*, where this arginine residue is at position 314 in the sequence. A R314L mutant showed a complete lack of activity.[21] In another kinetics study, results on a R314Q and a D312N mutant were reported.[22] It was suggested that either mutant may affect the step between citryl-coenzyme A formation and hydrolysis, and that it must involve both Arg314 and D312 acting in unison, as both mutations produce a very similar effect. This latter point makes sense, because Arg314 and Asp312 form a salt-bridge and can therefore be understood as a conformational change involving both. This fits with our hypothesis because, like Arg314, Asp312 is situated in the small domain, suggesting that the forces acting on Arg314 are transmitted to the rest of the small domain partly through their salt-bridge interaction, so helping to induce closure. Thus, without this interaction, closure of the small domain upon the large would be impaired.

Conclusions

Targeting and exploring simulations have been performed to assess whether the crystallographic closed domain conformation is accessible to the unliganded enzyme from the crystallographic open conformation and *vice versa*. The results presented here suggest that, even with a strong bias introduced, the crystallographic closed domain conformation cannot be reached from the crystallographic open conformation. Likewise, it would appear that the crystallographic open domain conformation is inaccessible from the crystallographic closed conformation. The closed conformation is unlikely to occur in the unliganded state, as the enzyme needs to open to release citrate. This is supported by all available crystal structures, which are all liganded in the active site when closed. It is likely, therefore, that the interaction with the products helps to recycle the enzyme back to the open conformation where they are finally able to escape. This means that the presence of a barrier between the open and closed conformations for the unliganded state is of particular relevance for the open conformation, because it is only the open conformation that is likely to be unliganded. Our results indicate that the source of this energy barrier is related to the shift of α-helix 328–341 along its own axis relative to the large

† http://srs/ebi.ac.uk

domain. In a sequential model of ligand binding and domain closure,[14] oxaloacetate binds first to the large domain. This would put it in a position to interact with Arg329 at the base of the helix such that it would pull this helix downwards in a direction that is roughly parallel with its own axis. Thus, it is proposed that the source of the energy barrier lies in the interactions of this helix with its parallel partner in the large domain and that in going from the open to closed domain conformation it is the interaction primarily with Arg329 that provides the energy to overcome this barrier. In the unliganded state, the domains are able to close or open partially but without the shift of this helix the full domain movement is unable to occur. Our results indicate that in citrate synthase the open unliganded enzyme will remain in a relatively open conformation ready to receive the substrate, thus increasing the efficiency of the enzyme over one that is closed part of the time, with its binding site inaccessible to the substrate. The substrate therefore is the key that is able to unlock the mechanism that prevents the domains to close fully. In this sense, the substrate catalyses the domain closure.

Methods

Molecular dynamics simulations

The details of the protocols used in the simulations performed in this work are as described.[7] In short, MD simulations were performed on fully solvated dimers (~80,000 atoms in total) of pig heart citrate synthase using GROMACS.[23] The initial structures were the crystallographic closed structure[15] liganded both to citrate, and coenzyme A (PDB accession code 2CTS), and the crystallographic open structure liganded to citrate, but bound differently from when it is a product (PDB accession code 1CTS).[15] None of these ligands was included in the simulations.

Essential dynamics sampling

The EDS technique is used to increase or decrease the distance from a reference conformation. This expansion or contraction is performed in a subspace (see below). At each time-frame, the usual MD step is performed and the distance in the subspace between the current conformation and the reference conformation is calculated. The step is accepted if this distance does not decrease in the case of expansion, or does not increase in the case of targeting. Otherwise, the coordinates and velocities are projected radially onto the hypersphere (in the subspace) centred on the reference conformation, with a radius given by the distance from the reference in the previous step (see Figure 8). Here, the sampling is performed in two distinct modes: targeting and exploring. In targeting, contraction is performed to a specified target conformation. In exploring, initial expansion occurs from a specified reference conformation (e.g. the crystallographic open conformation), but when expansion is halted according to two parameters described below, the final conformation becomes the new reference conformation from which a new expansion is started. There are two parameters required for the EDS in the exploring mode: the maximum number of sampling cycles before changing the origin of expansion, and the slope, which sets a minimum on the rate of expansion. These parameters were fixed to 5000 steps and 0.0004 nm/step, respectively. Targeting mode simulations were stopped when the radius failed to decrease any further in a number of consecutive steps. In all the simulations reported, EDS was applied to one monomer only, the other being allowed to undergo free MD.

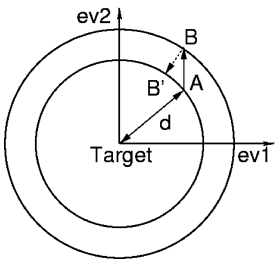

Figure 8. Essential dynamics sampling: example for the contraction procedure in a 2D case. (A) Structure at step i; (B) structure at step $i + 1$ generated by molecular dynamics; (B′) new structure at step $i + 1$ determined by the essential dynamic sampling procedure in the targeting mode. ev1 and ev2 represent eigenvectors 1 and 2, which determine the space within which the targeting is performed.

Sampling subspace

The sampling space was the six eigenvectors that represent the relative rigid-body motion of the two domains of a single monomer. The domains were assigned as before[7] and comprised a large domain of residues 1–55, 67–278 and 378–437, and a small domain of residues 56–66 and 278–377, which were determined by the program DynDom.[24,25] These six eigenvectors were determined as follows. The trajectories of both monomers from the open simulations in the previous work[7] were combined to give an equivalent single-monomer trajectory of 12 ns. This trajectory was used for a rigid-body essential dynamics analysis that was slightly different from that described previously. The external motion of the monomer was removed by superposing each conformation on the experimental open conformation. Superposition was done using the usual least-squares fitting procedure. Intradomain fluctuation was removed from this trajectory of intramonomer fluctuation by superposing the experimental open domain conformations on their respective domain conformations at each time-frame. Superposition was done using C^{α} atoms only. This gave a trajectory of the two domains as rigid bodies. Then conventional essential dynamics analysis (principal component analysis) was applied to this trajectory, resulting in six non-zero eigenvalues, which collectively quantify the amount of relative motion there is between the two domains. The eigenvectors corresponding to these six eigenvalues determine the subspace for the EDS. Applying EDS in this subspace allows us to encourage one domain to move relative to the other, and to explore domain conformations accessible from a specific conformation (exploring mode), or to target a specific domain conformation (targeting mode). Note that these constraints are applied to the domain conformations only, and all the intradomain degrees of freedom are allowed to undergo free MD.

Visualisation of relative motion of the domains

The trajectories are displayed by projecting onto the 2D space specified by the first two eigenvectors of the rigid-body essential dynamics analysis of the combined open trajectories from the previous work.[7] Roughly 80% of the domain fluctuation in the combined open trajectories occurred in this 2D space, and the domain movement between the two crystallographic conformations could be represented to 98% by these two modes. Thus, projecting the new trajectories from the EDS simulations onto this space enables one to visualise the relative motion of the domains.

DynDom and Dom_Select

In order to analyse the domain movements that occur, two programs have been used. The program DynDom[24] takes two conformations and determines dynamic domains, hinge axes and hinge-bending regions. It determines domains automatically on the basis of the conformational change itself. Here, we have used also an unreleased program Dom_Select that allows the user to specify the domains themselves by residue number ranges. Once the user specifies the domains, the hinge axis is determined in the same way as in the DynDom program.

Rigid-body RMSD

This quantity was used in the previous analysis.[7] Consider a part of the protein that moves from an initial to a final position. At the same time, the internal conformation of this part changes. The rigid-body movement of the part between the initial and final positions is calculated by superposing the initial conformation of the part on the final conformation. Thus, one has the initial conformation in two positions, the initial and final. The rigid-body root mean-square deviation, RG_RMSD, is simply derived from the displacement of each atom between these two positions.

Helix_Shift

The unreleased program Helix_Shift calculates the shift of an α-helix along its own axis. As above, consider a helix that moves from an initial to a final position. At the same time, the internal conformation of the helix changes slightly. The rigid-body movement of the helix between the initial and final positions is calculated by superposing the initial conformation of the helix on the final conformation of the helix. The movement of the helix along its own axis is then determined by calculating the distance between the centres of mass of the helix in these two positions and projecting this distance onto the helical axis of the helix in its initial position. The direction of this axis is estimated by superposing an ideal α-helix of identical length with its axis along the *z*-axis onto the real helix. The last column of the rotation matrix from this least-squares superposition gives the direction of the axis of the real helix from which the projected distance can be calculated.

Acknowledgements

I.D. thanks Professor Alfredo Di Nola for his support.

References

1. Schulz, G. E. (1991). Domain motions in proteins. *Curr. Opin. Struct. Biol.* **1**, 883–888.
2. Gerstein, M., Lesk, A. M. & Chothia, C. (1994). Structural mechanisms for domain movements in proteins. *Biochemistry*, **33**, 6739–6749.
3. Hayward, S. (1999). Structural principles governing domain motions in proteins. *Proteins: Struct. Funct. Genet.* **36**, 425–435.
4. Baker, E. N., Rumball, S. V. & Anderson, B. F. (1987). Transferrins—insights into structure and function from studies on lactoferrin. *Trends Biochem. Sci.* **12**, 350–353.
5. Millet, O., Hudson, R. P. & Kay, L. E. (2003). The energetic cost of domain reorientation in maltose-binding protein as studied by NMR and fluorescence spectroscopy. *Proc. Natl Acad. Sci. USA*, **100**, 12700–12705.
6. Liao, D. I., Karpusas, M. & Remington, S. J. (1991). Crystal structure of an open conformation of citrate synthase from chicken heart at 2.8 Å resolution. *Biochemistry*, **30**, 6031–6036.
7. Roccatano, D., Mark, A. E. & Hayward, S. (2001). Investigation of the mechanism of domain closure in citrate synthase by molecular dynamics simulation. *J. Mol. Biol.* **310**, 1039–1053.
8. Wiegand, G. & Remington, S. J. (1986). Citrate synthase, structure, control, and mechanism. *Annu. Rev. Biophys. Biophys. Chem.* **15**, 97–117.
9. Amadei, A., Linssen, A. B. M., de Groot, B. L., van Aalten, D. M. & Berendsen, H. J. C. (1996). An efficient method for sampling the essential subspace of proteins. *J. Biomol. Struct. Dynam.* **13**, 615–625.
10. de Groot, B. L., Amadei, A., Scheek, R. M., van Nuland, N. A. & Berendsen, H. J. C. (1996). An extended sampling of the configurational space of HPr from *E. coli*. *Proteins: Struct. Funct. Genet.* **26**, 314–322.
11. de Groot, B. L., Amadei, A., van Aalten, D. M. & Berendsen, H. J. C. (1996). Towards an exhaustive sampling of the configurational space of the two forms of peptide hormone guanylin. *J. Biomol. Struct. Dynam.* **13**, 741–751.
12. Roccatano, D., Daidone, I., Ceruso, M.-A., Bossa, C. & Di Nola, A. (2003). Selective excitation of native fluctuations during thermal unfolding simulations: horse heart cytochrome *c* as a case study. *Biophys. J.* **84**, 1876–1883.
13. Daidone, I., Amadei, A., Roccatano, D. & di Nola, A. (2003). Molecular dynamics simulation of protein folding by essential dynamics sampling: folding landscape of horse heart cytochrome *c*. *Biophys. J.* **85**, 2865–2871.
14. Hayward, S. (2004). Identification of specific interactions that drive ligand-induced closure in five enzymes with classic domain movements. *J. Mol. Biol.* **339**. In the press.
15. Remington, S., Wiegand, G. & Huber, R. (1982). Crystallographic refinement and atomic models of two different forms of citrate synthase at 2.7 and 1.7 angstoms resolution. *J. Mol. Biol.* **158**, 111–152.

16. Lee, R. A., Razaz, M. & Hayward, S. (2003). The DynDom database of protein domain motions. *Bioinformatics*, **19**, 1290–1291.
17. Bairoch, A. (2000). The ENZYME database in 2000. *Nucl. Acids Res.* **28**, 304–305.
18. Boeckmann, B., Bairoch, A., Apweiler, R., Blatter, M. C., Estreicher, A., Gasteiger, E. *et al.* (2003). The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucl. Acids Res.* **31**, 365–370.
19. Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucl. Acids Res.* **22**, 4673–4680.
20. Hulo, N., Sigrist, C. J. A., Le Saux, V., Langendijk-Genevaux, P. S., Bordoli, L., Gattiker, A. *et al.* (2004). Recent improvements to the PROSITE database. *Nucl. Acids Res.* **32**, D134–D137.
21. Anderson, D. H. & Duckworth, H. W. (1989). Mutation of amino-acids thought to polarize the oxaloacetate carbonyl in citrate synthase severely reduces but does not abolish activity of the enzyme. *Biochem. Cell Biol.—Biochim. Biol. Cell*, **67**, 98–102.
22. Man, W. J., Li, Y., O'Connor, C. D. & Wilton, D. C. (1994). The effect of replacing the conserved active-site residues His-264, Asp-312 and Arg-314 on the binding and catalytic properties of *Escherichia coli* citrate synthase. *Biochem. J.* **300**, 765–770.
23. Berendsen, H. J. C., van der Spoel, D. & van Drunen, R. (1995). GROMACS: a message passing parallel molecular dynamics implementation. *Comput. Phys. Commun.* **91**, 43–56.
24. Hayward, S. & Berendsen, H. J. C. (1998). Systematic analysis of domain motions in proteins from conformational change: new results on citrate synthase and T4 lysozyme. *Proteins: Struct. Funct. Genet.* **30**, 144–154.
25. Hayward, S. & Lee, R. A. (2002). Improvements in the analysis of domain motions in proteins from conformational change: DynDom version 1.50. *J. Mol. Graph. Model.* **21**, 181–183.
26. Sayle, R. & Milner-White, E. J. (1995). Rasmol: biomolecular graphics for all. *Trends Biochem. Sci.* **20**, 374–375.
27. Kraulis, P. J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallog.* **24**, 946.
28. Merritt, E. A. & Murphy, M. E. P. (1994). Raster3d version-2.0—a program for photorealistic molecular graphics. *Acta Crystallog. sect. D*, **50**, 869–873.

Edited by R. Huber

(Received 22 December 2003; received in revised form 18 March 2004; accepted 5 April 2004)