JCTC Journal of Chemical Theory and Computation

Investigating the Conformational Preferences of Transforming Growth Factor-β Isoforms using Targeted Molecular Dynamics Simulations

Daniel J. Warner,*,† Ian C. Paterson,‡ and Richard B. Sessions§

Department of Medicinal Chemistry, AstraZeneca R&D Charnwood, Bakewell Road, Loughborough, Leicestershire, LE11 5RH, U.K., Department of Oral and Dental Science, Bristol Dental Hospital, University of Bristol, Lower Maudlin Street, Bristol, BS1 2LY, U.K., and Department of Biochemistry, School of Medical Sciences, University of Bristol, University Walk, Clifton, Bristol, BS8 1TD, U.K.

Received November 9, 2008

Abstract: We study the conformational preferences and mechanical properties of two isoforms of the cytokine transforming growth factor-β (-β1 and -β3) with atomistic detail and including the effects of explicit water. Targeted molecular dynamics simulations are used to perturb experimental "closed" conformations of both proteins into an "open" conformation, thus far only observed crystallographically for one of the two isoforms. The artificial restraints imposed by the protocol are later released, allowing the two covalently bound units of each homodimer to relax. Homology models of the two proteins are also constructed using the other as a template; models that are later subjected to the same process of perturbation into the open conformation and relaxation. On release, both simulations of transforming growth factor-β1 show a tendency to snap back toward the closed conformation, while those of transforming growth factor-β3 remain open for the remainder of the simulation, apparently consistent with measurements from a variety of experimental sources. Duplication of the simulations affords confidence that this observation reflects a genuine effect of the sequence, as opposed to an artifact of the conformations selected at the outset. The study provides a previously unseen level of detail, describing the structural and dynamic behavior of these proteins in solution, and brings us a step closer to understanding the complex relationship between sequence, structure, and signaling in this family of cytokines.

Introduction

TGF-β Signaling. Transforming growth factor-β (TGF-β) describes a family of multifunctional cytokines responsible for regulating a wide range of cellular processes including growth and differentiation.[1,2] Three mammalian isoforms of TGF-β have been identified (TGF-β1, -β2, and -β3), which exhibit a high degree of sequence identity (in the range of 70−76%). Despite this similarity, the differences that exist across the three isoforms are strictly conserved across species, reflecting each isoform's distinct biological properties, and hence their resistance to evolutionary pressures.[3,4] The type-I and type-II signaling receptors for TGF-β (TβR-I and TβR-II) each contain an extracellular ligand-binding, transmembrane, and intracellular Ser/Thr kinase domain.[5] Initially, TGF-β binds to the extracellular domain of TβR-II, forming a complex that is then able to recruit the second, type-I receptor.[6] Association of these two extracellular domains by TGF-β prompts transphorylation of the TβR-I kinase domain, which instigates a series of downstream signaling events.[5]

Structural Biology. X-ray crystal structures have been determined for both TGF-β2[7,8] and -β3[9] in isolation, whereas the structural information for TGF-β1 is derived from NMR spectroscopy.[10,11] These studies show all three

* To whom correspondence should be addressed. E-mail: dan.warner@astrazeneca.com.

† Department of Medicinal Chemistry, AstraZeneca R&D Charnwood.

‡ Bristol Dental Hospital, University of Bristol.

§ School of Medical Sciences, University of Bristol.

10.1021/ct800483q CCC: $40.75
Published on Web 02/11/2009

isoforms to associate in the form of a homodimer, with the two monomers cross-linked by a single disulphide bond (Cys77−Cys77). The structures are typically described as resembling a pair of hands clasped together, with a collection of β-strands representing the fingers, while α-helices H1 and H3 from each monomer represent the thumbs and wrists, respectively.

Despite such high similarity between TGF-β1 and -β3 in terms of their sequences and experimentally determined conformations, their structures and dynamic behavior in solution show marked differences. ^{15}N NMR relaxation studies indicate that the sections of the sequence encompassing α-helices H1 and H3 display much greater mobility in TGF-β3 than they do in -β1.[12] Notably, α-helix H3 is made up of the least well conserved residues across the TGF-β isoforms, indicating that this is the region of the protein where the biological properties are most likely to be modulated. The structures of TGF-β in isolation appear to support this concept, in that despite having the same space group and arrangement within each unit cell, the residues comprising H1 have notably higher thermal B-factors in 1TGJ (TGF-β3)[9] than in the two crystallographic structures of TGF-β2.[7,8] The observed flexibility of H3 is a more surprising finding, as in the crystalline state this region represents one of the most rigid parts of the molecule. However, the restraint of this helix in TGF-β3 is thought to be an artifact of the crystal packing because a number of residues are involved in intermolecular interactions in the crystal lattice.[9]

Further evidence for the disordering of α-helices H1 and H3 of TGF-β3 in solution is provided from circular dichroism experiments.[13] This study shows the helical content of TGF-β3 in solution to be as low as 4%, far lower than would be expected from the crystal structures of either this isoform or TGF-β2. The authors conclude by stating that, in contrast to TGF-β2, -β3 has the ability to adopt two distinct conformations. Insight into the possible nature of this second conformation has, at least in part, become available with the publication of the crystal structure of TGF-β3 in complex with the extracellular domain of TβR-II.[14] In this structure, the two monomers undergo significant reorientation with respect to each other, with one subunit effectively rotating by 101° with respect to the other. As might be expected, this "open" structure contains far lower helical content than any of the previously determined "closed" structures of TGF-β in isolation because the α-helices that comprise the thumb and wrist are considerably more mobile. This flexibility brings the thermal B-factors more closely into line with the exaggerated mobilities observed in the NMR experiments, making the open conformation a more reliable template for models of TGF-β3 in solution.

A significant advance in understanding the relationship between structure and function was recently made with the solution, to a resolution of 3.0 Å, the structure of the complete TGF-β:TβR-I:TβR-II complex.[15] Crucially, this complex indicates that TGF-β must adopt a closed conformation to recruit the extracellular domains of its type-I and type-II receptors and thus instigate the downstream signaling cascade. Prior to this work, it had only been possible to construct putative models of this complex,[6,16] based on combination of the TGF-β3:TβR-II complex structure[14] with that of the closely related bone morphogenetic protein 2 (BMP-2) in complex with the extracellular domain of the BMP receptor 1A (BR1A).[17]

Methodology Selection. The striking differences between the TGF-β isoforms, particularly in terms of their structural and dynamic behavior in solution, make this group of proteins compelling subjects for studies by molecular dynamics (MD) simulation. We have used targeted molecular dynamics (TMD)[18,19] to promote the transition from the closed form of TGF-β3 to the open form to produce a model that is more representative of this protein's behavior in solution.

Our reasons for adopting this approach are essentially 2-fold. First, while the actual number of residues that remain unresolved in the TGF-β3:TβR-II complex structure is fairly low, those sections of the sequence neighboring these residues appear frayed when compared to structures of the ligand alone and feature high crystallographic B-factors.[14] This raises a number of questions for anyone wishing to construct a homology model of the open conformation directly, in that it is unclear how far back to prune these termini to construct the missing fragment, and this potentially requires construction of a loop longer than would be recommended for most modeling packages. The second argument centers on the nature of the fragment to be inserted. The relative orientation of the monomers in the complex structure prohibits the introduction of a wrist helix as observed in the closed conformation, while construction of an extended loop would generate a starting point for a simulation with little chance of regaining this secondary structure, regardless of the shape of the conformational energy landscape.

As a point of comparison, we also performed the same experiment based on an experimental structure of TGF-β1, for which there is currently no evidence of such a second, partially denatured conformation existing. The conformational preferences and secondary structure composition of these two isoforms are presented, and their implications for recognition and signaling are discussed.

Methods

System Preparation. The minimized average NMR structure of TGF-β1[10] and X-ray crystal structure of TGF-β3[9] were downloaded from the RCSB Protein Data Bank[20] (entries 1KLC and 1TGJ). The 224 (2 × 112) residue dimers were parametrized within the leap module of AMBER 8[21] using the ff03 force field,[22] with disulphide bonds connecting the cystine residues at positions 7, 15, 44, and 48 with those at 16, 78, 109, and 111 in each monomer, respectively. The two remaining cystine residues at position 77 formed the only covalent bond connecting the two dimers. The histidine residues at 34, 40, and 68 in TGF-β1 and those at 34 and 58 in TGF-β3 were all designated as ε tautomers. Default ionization states were used for all other residues.

Homology models of the two proteins were also constructed on the basis of the structure of the alternate isoform,

that is, a model of TGF-β3 was based on the 1KLC structure and that of TGF-β1 was based on the 1TGJ structure. This task was performed manually; atoms common to any mutated residues were preserved and renamed where necessary, while the remainder of the side-chains had their coordinates deleted. The leap module was allowed to complete any remaining atoms according to residue templates, and the systems were parametrized as above. All four structures were immersed in rectangular boxes containing over 8000 TIP3P[23] water molecules. Neutralization required the addition of 8 Cl^- ions to each of the TGF-β1 systems and 2 Na^+ ions to each of the TGF-β3 systems. The structures were allowed to relax via 50 steps of steepest descent, followed by conjugate gradient minimization until the energy gradient had been reduced to 0.5 kcal mol^{-1}.

Target structures representing the open conformation were based upon the X-ray crystal structure of the TGF-β3:TβR-II complex (PDB entry 1KTZ).[14] Because the structure of TGF-β3 in the complex is incomplete and to minimize any potential bias (see discussion), monomers from each of the systems above were superimposed onto the complex structure, creating a total of four target conformations (Figure 1). Those residues where the tertiary structure remained unchanged between the open and closed conformations were then identified (resides 18−44 and 81−108) and designated as the residues to which the TMD restraints would be applied (α-carbon atoms only).

Molecular Dynamics Protocol. All simulations were performed with a 2 fs time step and with the use of SHAKE to constrain all bonds to hydrogen at equilibrium values.[24] Electrostatics beyond a 12 Å cutoff were calculated using particle mesh Ewald.[25] In the initial phase, simulations were performed at constant temperature and volume (NVT ensemble), with the temperature regulated using Langevin dynamics and a collision frequency of 5 collisions ps^{-1}; 100 ps simulations were performed on the solvent and counterions at 100 K to stabilize interactions with the protein. The water was heated to 300 K over a further 100 ps, with a harmonic restraint of 100 kcal mol^{-1} $Å^{-2}$ to keep the protein atoms close to the original structures. These restraints were then gradually reduced over a further 600 ps, before allowing the simulations to run unrestrained for a period of 200 ps.

At this point, the protocol was switched to the NPT ensemble to allow equilibration of the density, and the temperature was coupled to an external bath using a weak coupling algorithm.[26] Temperature and pressure coupling was performed at 1 ps intervals, with isotropic position scaling ensuring the pressure remained constant at 1 bar. The simulations were allowed to run in this state for a period of 10 ns.

Following this period of unrestrained simulation of the closed conformations, restraints were applied to the structures to induce the open state. At each step, the current simulation structure was superimposed onto the target using the α-carbon atoms of residues 18−44 and 81−108 from both monomers, and the rmsd was calculated between the two sets of atomic coordinates. The total energy penalty for deviation from the target rmsd was set to 5 kcal mol^{-1} $Å^{-2}$ $atom^{-1}$ (550 kcal mol^{-1} $Å^{-2}$). Initially, the target rmsd was set to the rmsd at 11 ns (typically in the region of 7.5−8 Å) and then incrementally reduced over a period of 5 ns to a value of 2 Å. Finally, the structures were allowed to relax, through gentle reduction of the restraint over a period of 1 ns, before allowing the simulations to run unrestrained for a further 33 ns. Overall, deposition of coordinates at 100 ps intervals generated a total of 500 structures for analysis.

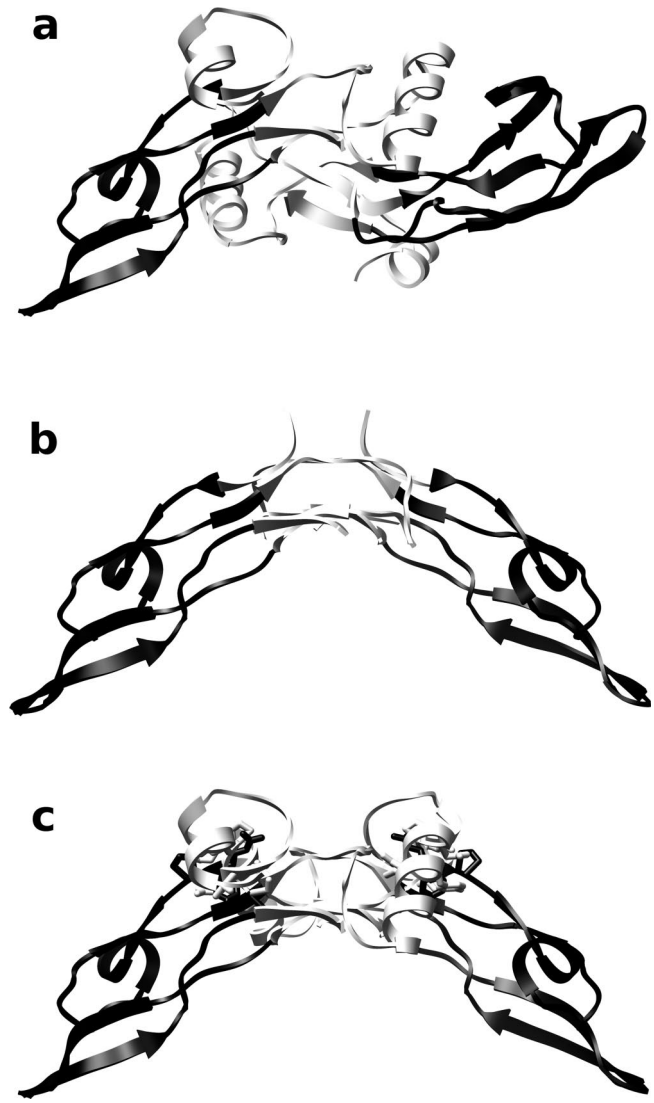

Figure 1. Construction of a model for the open conformation of TGF-β. (a) Crystal structure of TGF-β3 in isolation depicting the closed conformation.[9] (b) Crystal structure of TGF-β3 in complex with the extracellular domain of TβR-II (receptor not shown).[14] (c) Superimposition of monomers from a onto b to form a template for the open conformation. Residues exhibiting minimal displacement between b and c are rendered in black and were used as the α-carbon atoms to which the TMD restraints were applied (resides 18−44 and 81−108). Stick rendering highlights the side-chains involved in intermonomer steric overlap.

Trajectory Analysis and Figure Generation. The rmsd calculations, secondary structure, and native contact analyses were all performed using the AMBER module ptraj. Native contacts were defined as any pairwise atomic contacts (<5 Å) present in the initial structure (closed). In each system, contacts were identified within two individual subsets of residues comprising α-helix H3 of one monomer (residues

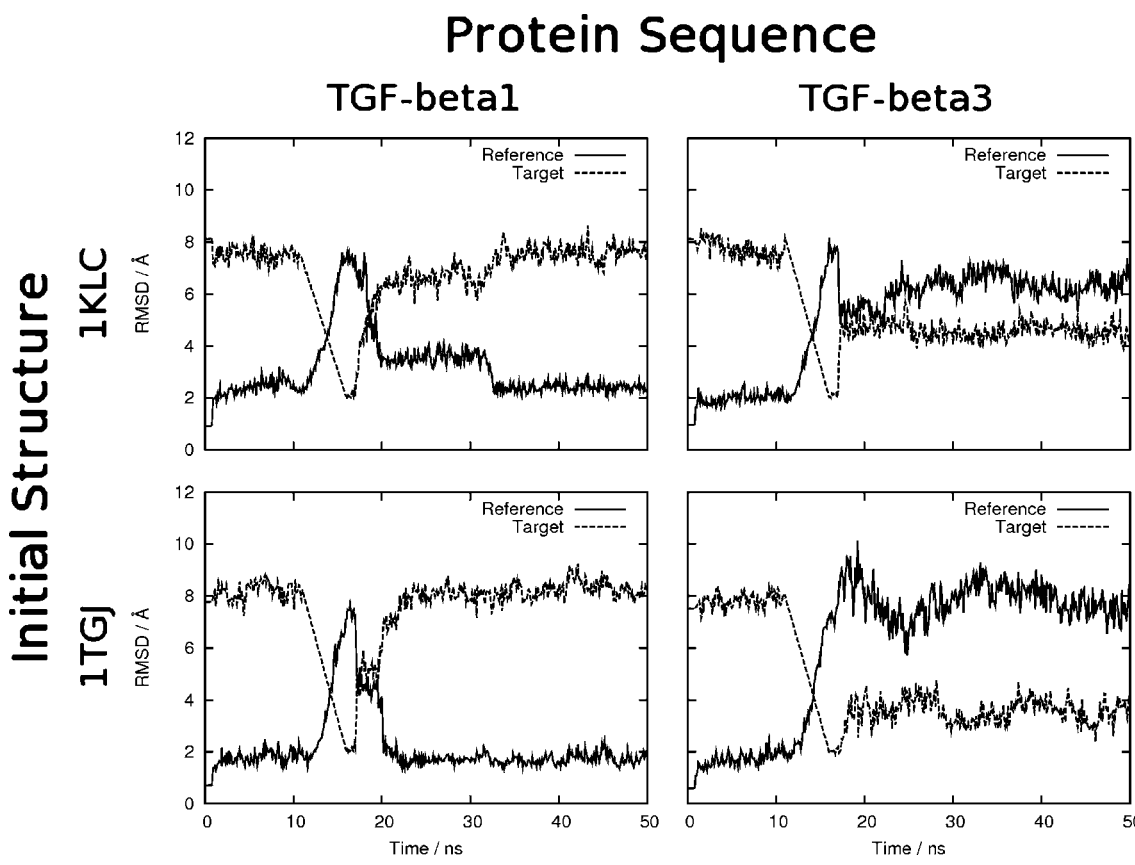

Figure 2. rmsd calculations for α-carbon atoms, tracking the conformational transition from the free, closed state to the restrained, open state and eventual release in the four simulations. Solid lines show comparisons of the structures to the closed, initial structure, calculated for all residues. The dotted lines show comparisons to the open, target structure, calculated only for those residues to which the restraint was applied (residues 18−44 and 81−108).

57−68 inclusive) and the full opposite monomer. The presence of such contacts was evaluated for each trajectory structure and averaged over the periods prior to and following the TMD restraints being applied (1−11 and 17−50 ns, respectively).

Computational glycine scanning was performed using the mm_pbsa perl scripts provided in the AMBER package. This method involves mutation of each residue to glycine through the removal all side-chain atoms except C_β, which is replaced by a hydrogen atom. Using this technique, it is possible to quickly estimate the effect of mutating each residue to glycine from a single trajectory. The GBSA method was used to evaluate the electrostatic and nonpolar components of the free energy of solvation,[27] thus the free energy difference between the wild-type protein, G_{wt}, and the mutant, G_{mut}, was represented by

$$\Delta G = G_{wt} - G_{mut}$$

where

$$G = E_{int} + E_{vdW} + E_{ele} + G_{sol} - TS$$

and

$$G_{sol} = G_{GB} + \gamma_{SASA} + \beta$$

In this study, we assume that the difference in the overall entropy between the wild-type protein and the mutant is negligible, and so no attempt was made to calculate *TS*. In the calculation of G_{GB}, a dielectric constant of 1 was used for the interior of the protein, and 80 was used for the surrounding solvent; γ represents the surface tension applied to the solvent accessible surface area (SASA, calculated by the LCPO method[28]) and was set to 0.0072 kcal mol^{-1} Å^{-2}, with the offset, β set to 0 kcal mol^{-1}. Finally, the free energy differences were mean-centered on a per-residue basis to facilitate the identification of energetic hot-spots along the trajectory. Proline and disulphide bonded cystine residues were omitted from the analysis.

Molecular graphics images were produced using the UCSF Chimera[29] package from the Computer Graphics Laboratory, University of California, San Francisco. All plots were produced using Gnuplot 4.0.

Results

Backbone Structural Analyses. To assess the stability of the simulations, rmsd calculations were performed on the α-carbon atoms, comparing the original experimental structures with those generated during the course of the simulation (Figure 2). It can be seen that for a little over 10 ns from the outset, the simulated structures maintain an average deviation of around 2 Å from the initial conformation in all four systems. This provides evidence that the unrestrained simulations of the closed conformation were stable and, notably, that this was equally the case for the homology models as for the experimental structures. During this time, none of the four systems gave any indication of spontaneously adopting the open conformation.

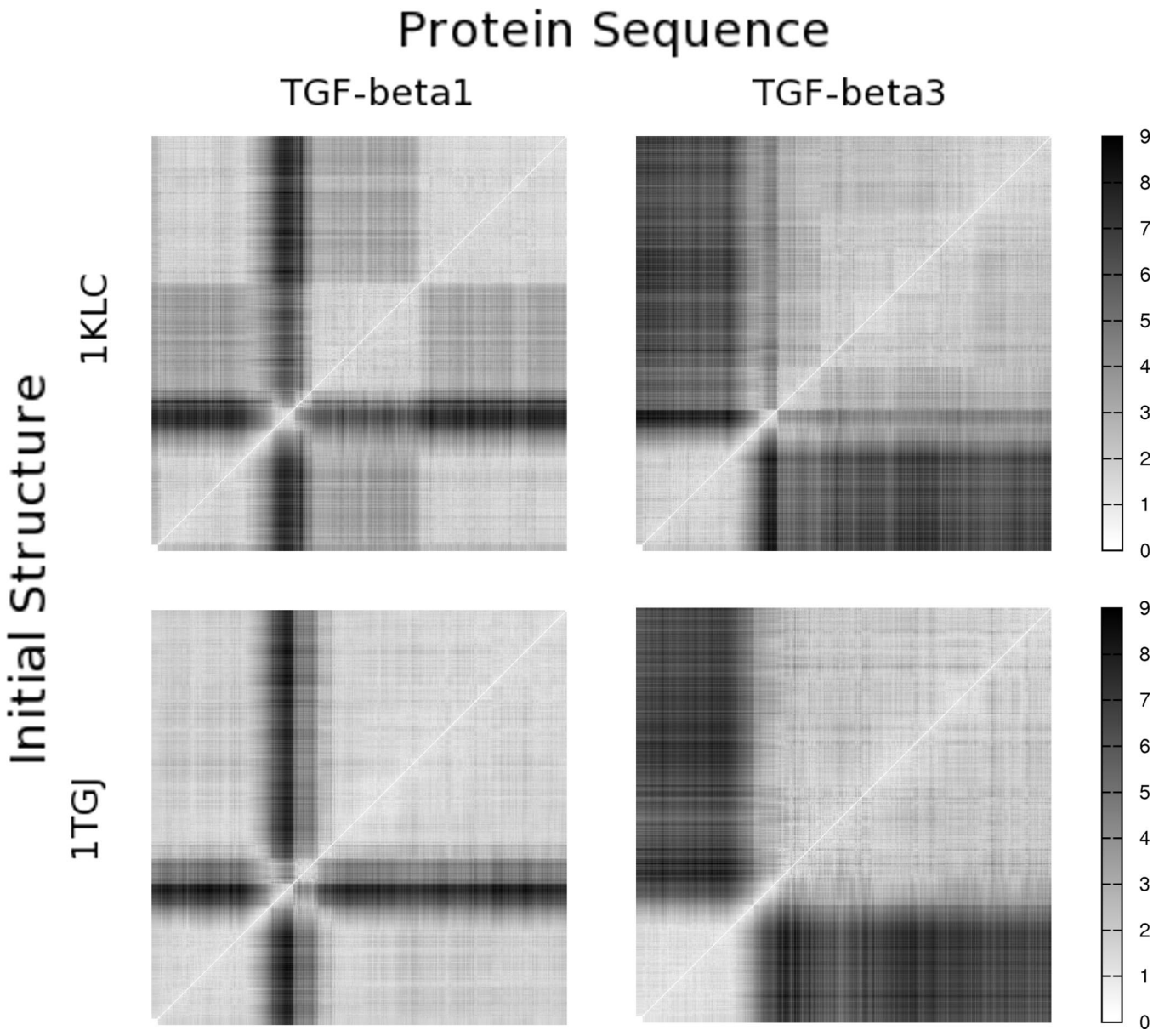

Figure 3. 2D rmsd plots comparing each snapshot with every other snapshot within a given simulation. Low rmsd values are represented as white squares, while high values are black. Simulation time begins at the bottom/left of the plots and ends at the top/right. Calculations were based on the α-carbon atoms of all residues.

From 11 ns, the rmsd between the simulated structures and the original (solid line) rises sharply as the TMD restraints are applied, and the relative orientation of the dimers is forced away from the closed state, toward the open conformation. Conversely, when comparisons are made with the target structure (dotted line), the rmsd falls as the target value is reduced, with the measured rmsd eventually reaching around 2 Å at a time of 16 ns.

It is from this point forward, as the structures are allowed to relax, that the behavior of the four simulations differs most significantly. In the two simulations of TGF-β1, the rmsd from the initial structure begins to fall as the restraints are released, indicating that the structures are moving back toward the original, closed conformation. By the end of the simulations the α-carbon atoms are, on average, around 2 Å from where they began, and exhibit a greater degree of similarity to the closed conformation than they do the target. In contrast to this, the TGF-β3 simulations showed no such tendency to revert back to the closed state, eventually finishing with rmsd values of 6.5 and 8.2 Å from their initial conformations. Both the simulation that started from the TGF-β3 crystal structure, and that which started from a homology model based on the TGF-β1 NMR structure, ended with structures with a greater similarity to the open target (with an rmsd of approximately 4 Å in both cases) than to the initial, closed conformation.

The 2D rmsd plots (Figure 3) provide a clear, visual means of identifying the proteins' transitions between different regions of conformational space. The clearly defined white squares in the bottom left corners of the plots represent the early part of the simulations, where all four systems occupied the clearly defined closed state. The large proportion of off-diagonal white points in this area shows that the structures

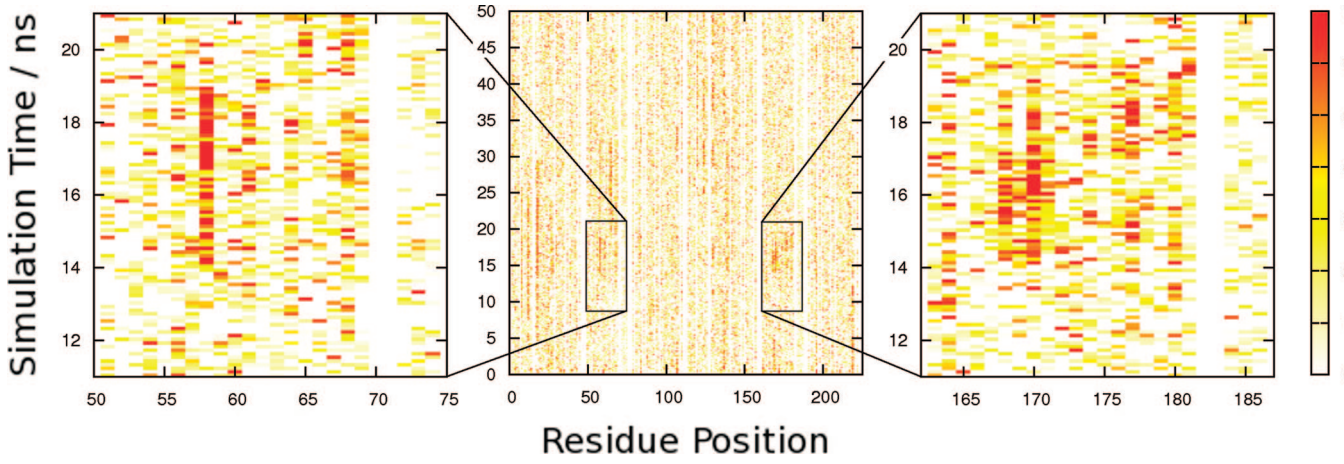

Figure 4. Computational glycine scanning applied to the simulation of TGF-β1 (taking the NMR structure,[10] 1KLC as the initial conformation). Points on the maps represent the free energy difference, ΔG between the wild-type and glycine mutant for each snapshot in the simulation, mean-centered for each residue. White points represent stable residues, where this value is less than 0 kcal mol^{-1}, and red points represent unstable residues, where this value exceeds 7 kcal mol^{-1}.

in the first 11 ns of any given simulation were comparatively very similar to each other.

As the restraints are applied and the conformations are encouraged to move toward the open state, the calculated deviations from the early structures begin to increase sharply, and consequently, the off-diagonal elements of the matrices become much darker in all four systems.

Once again, it is from this point on, when the TMD restraint is released, that the four simulations begin to differ. In the final phase of the simulations, it can be seen that the two simulations of TGF-β1 contain much lighter off-diagonal elements than do the simulations of TGF-β3. These lighter areas indicate the exploration of conformational space in the latter stages of the simulations that bears a high similarity to the initial, closed state. The apparent pattern of a dark cross against a light background is characteristic of the A $\rightarrow$ B $\rightarrow$ A state behavior. In contrast to this, the TGF-β3 simulations simply continue to explore space surrounding the twisted, open conformation, without returning to the experimental structures. It is this simple A $\rightarrow$ B state behavior that results in the apparent division of the 2D rmsd plots into four distinct corners.

Changes to the secondary structure content of the proteins during the course of the simulations were also calculated. From the outset, all four simulations had a much greater proportion of β-strands than α-helices, with a relative composition of around 40 and 15%, respectively. This composition is subject to a degree of fluctuation as the simulations progress, and while there is a general tendency for the strand content to deteriorate slightly in all simulations, there is no direct link between this variation and the application or release of the restraints. Ultimately, there is little change in the secondary structure of any of the proteins throughout the whole process, with all 4 systems retaining about 15% α-helix and 35% β-sheet at 50 ns.

Computational Glycine Scanning. The results from computational glycine scanning reveal the location of hot-spots in terms of the free energy difference between the wild-type and mutant protein (Figure 4). The red points indicate the location of structures where ΔG is high, that is to say, where the wild-type protein is considerably less stable than the glycine mutant. Particularly interesting in this study are those residues where the energy appears to spike at time points in the range 11−16 ns, because this is the time when TGF-β1 was forced to adopt the open conformation. The area of the plot where this effect appears most prevalent is in the region of residues 164−180 (Figure 4, right) and, to a lesser extent, 52−68 (Figure 4, left). These regions of the protein encompass the wrist helix of each monomer, which undergoes significant displacement during the course of the conformational transition. The most noticeable degree of instability affects residue 58 (residue 170 in monomer B), a buried tyrosine residue that undergoes mutation to histidine in TGF-β3.

Native Contact Analysis. A key feature of how the TGF-β3 structure is able to adapt to the open conformation is a change in the relative position of α-helix H3 in one of the two monomers. This translation avoids clashes with the opposite monomer, while maintaining native contacts at the dimer interface. To explain and quantify this effect, we tracked the presence of native atomic contacts within H3 and its opposite monomer over two periods, *pre* and *post* application of the TMD restraint. The results of this analysis are presented in Table 1.

As a result of reverting to the closed conformation, the two simulations of TGF-β1 show little difference in the prevalence of native atomic contacts before and after application of the restraints. In all four, such H3 plus monomer subsets, the loss of native contacts is in the region of just 1−2%. In contrast to this, the two TGF-β3 simulations both remained in the open state, and as would be expected, a significant proportion of native (closed state) contacts are lost. It appears from the data that in both simulations, the loss of native contacts is more marked in one monomer than in the other, that is, one α-helix shifts to accommodate the opposite monomer, thus retaining a number of closed state contacts at the dimer interface, while the alternate subset makes 8−10% less native contacts than it did in the closed conformation. The same effect is also apparent from the visual representation of the final structures presented in

Table 1. Number of Native Atomic Contacts As a Percentage of the Number Present in the Initial Structure[a]

initial structure	protein sequence	residue subset	pre-TMD	post-TMD	difference
1KLC	TGF-β1	H3 (monomer A) + monomer B	74.7%	73.1%	−1.6%
1KLC	TGF-β1	H3 (monomer B) + monomer A	74.2%	73.2%	−1.0%
1TGJ	TGF-β3	H3 (monomer A) + monomer B	81.6%	73.3%	**−8.3%**
1TGJ	TGF-β3	H3 (monomer B) + monomer A	80.5%	76.8%	−3.6%
1KLC	TGF-β3	H3 (monomer A) + monomer B	79.1%	76.2%	−2.9%
1KLC	TGF-β3	H3 (monomer B) + monomer A	79.8%	69.9%	**−9.9%**
1TGJ	TGF-β1	H3 (monomer A) + monomer B	81.7%	80.2%	−1.6%
1TGJ	TGF-β1	H3 (monomer B) + monomer A	82.2%	80.8%	−1.4%

[a] Values are quoted for the periods pre- and post-application and release of the TMD restraints (1−11 and 17−50 ns, respectively). The difference column quotes the loss of native contacts as a result of the conformational transition.

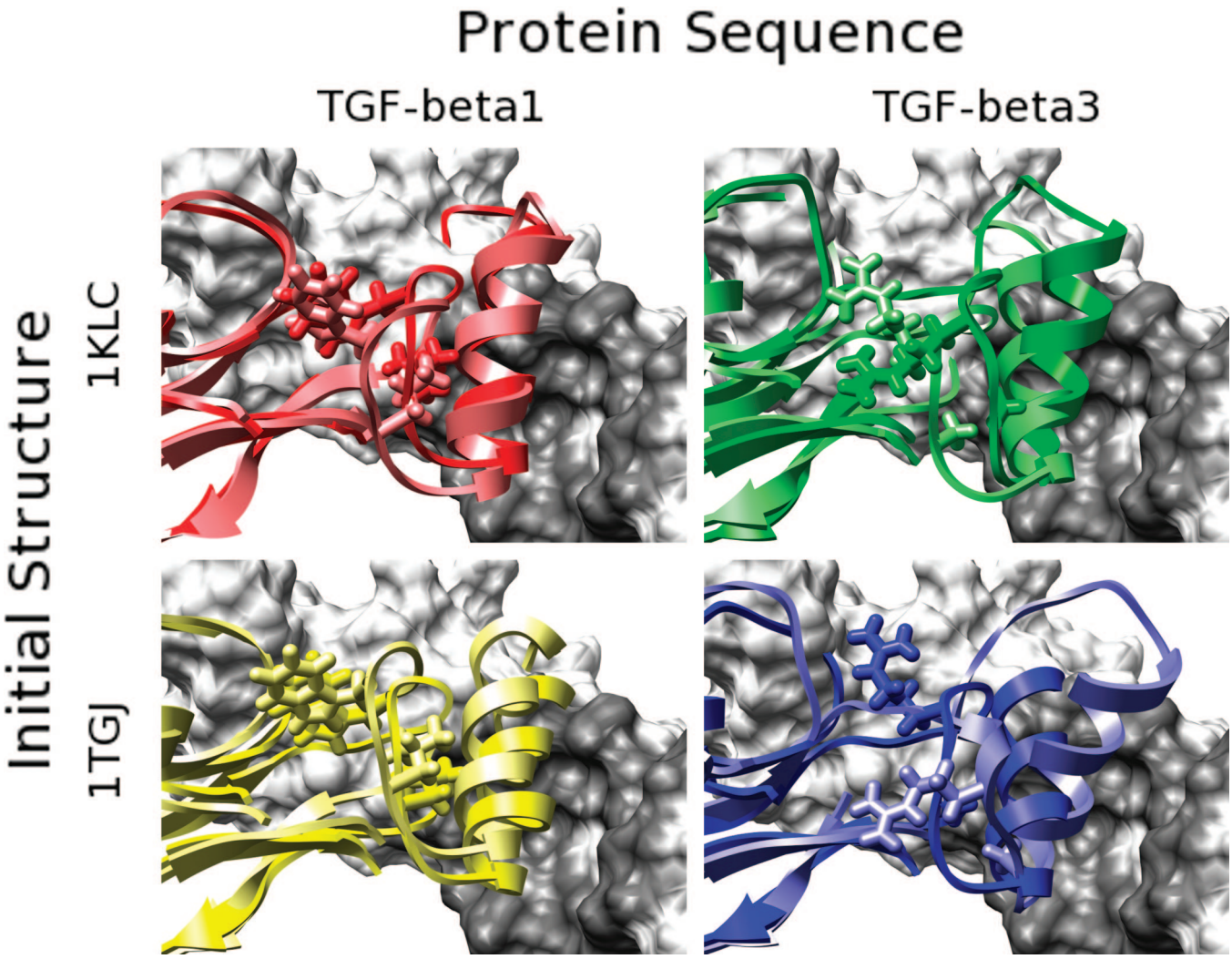

Figure 5. Cartoon representations of all four final structures following 50 ns of simulation. Monomer A in each system is shown in bold color, while monomer B is represented in a lighter shade. All 8 monomers are presented in the same frame of reference by overlaying each structure onto the final coordinates of TGF-β3 monomer B (1TGJ simulation). Monomer A from the same simulation is represented as a white surface, with those residues which form native contacts with α-helix H3 rendered in dark gray. Mutated residues in the loop preceding the helix are represented as sticks (residues 52 and 54).

Figure 5. In the TGF-β3 simulation that began from 1TGJ, it is monomer B (light blue cartoon) that shifts to regain contact with the concave groove in the opposite monomer (dark gray surface), while in the simulation that started from the 1KLC structure, it is monomer A (bold green cartoon) that undergoes this same transformation.

Along with the observed shifts of α-helix H3 in TGF-β3, we also see concerted changes in the conformation of the preceding loop, which encompasses residues Arg 52 and Ala 54. In both simulations, a change in the rotamer state of Arg 52 appears to be involved, featuring loss of the crystallographic interaction with Glu 12, and an approximate doubling in the population of the first solvation sphere (number of solvent residues <3.4 Å). The loss of the interaction with Glu 12 has the effect of releasing the tension in this loop, thus allowing the entire helix to change conformation. Notably in TGF-β1, where residues 52 and 54 are mutated to tryptophan and leucine, respectively,

solvent exposure of Trp 52 remains at pretransition levels, and the loop undergoes no such conformational change.

Discussion

Simulation Stability. The initial phase of the 1D rmsd plots allows comparison of the stability of the simulations, not only between one isoform and the other, but also between the simulations that began from an experimental structure of a particular isoform, and the homology model that was built using the other as a template. One may have expected that simulation of the solution phase NMR structure of TGF-β1 would generate structures in closer agreement with the experimental structure than the simulation of TGF-β3 because of the need for TGF-β3 to overcome any crystal packing artifacts present in the structure and adapt to the presence of aqueous solvent. This proved not to be the case. An encouraging result is that the same level of deviation from the experimental structures is exhibited in the simulations of the homology models. This indicates that the process of building the models did not introduce any serious clashes in the system that could not be rectified through minimization or the carefully controlled warm-up process, presumably because of the high level of sequence similarity across the two isoforms.

Conformational Preferences of TGF-β Isoforms. The major differences in the two isoforms' structures and behavior only became apparent following the release of the TMD restraints. When the perturbed structures were presented with an opportunity to relax, both simulations of TGF-β1 immediately moved away from the induced open conformation and began to revert back toward the initial, closed state. This observation was in stark contrast to the two simulations of the other isoform studied, TGF-β3, which remained comparatively comfortably in the open state.

These observations indicate that the process of adopting the open conformation is energetically far less far favorable for TGF-β1 than it is for TGF-β3. It is particularly encouraging to see the conformations favored by the homology models mimicking those of the equivalent isoform, rather than the isoforms on which the models were based. This allows confidence that the observations are a direct result of the protein sequence and not simply an artifact stemming from the fact that the simulations were initiated from different conformations. The results are in excellent agreement with the current experimental evidence surrounding these two isoforms. After all, it was an X-ray crystal structure of TGF-β3 in complex with the type II TGF-β receptor that was used as the initial template for the open conformation of the homodimer. TGF-β1, on the other hand, has only ever been observed in the closed state.

It should be noted that selection of a TGF-β3 structure for the TMD template had the potential to introduce a bias into the experiment. That is to say, if the systems had been encouraged to adopt the precise backbone conformation of TGF-β3, simulations of other isoforms would be considerably destabilized regardless of the relative orientation of the monomers. For example, Thr 87 in TGF-β3 is substituted for proline in TGF-β1, a residue which is far more limited in terms of main-chain flexibility. This residue may have encountered serious difficulties in adopting the backbone conformation of the corresponding threonine in the template. However, such bias was avoided by superimposing each individual monomer from the closed structures onto the complex at the time of target construction, the result being that the backbone conformation of the target then became allied to the initial experimental structure. Consequentially, we can confidently attribute the preferences exhibited by the two isoforms of TGF-β to the nature of the intramolecular interactions within the protein.

Dependence of Secondary Structure on Monomer Orientation. The relative orientation of the monomers in the open conformation poses a direct conflict with the location of α-helix H3 as exhibited in the closed crystal structure (Figure 1). For this reason, the structural stability of this section is thought to be crucial in determining the ability of TGF-β isoforms to adopt the open conformation. In particular, it has been suggested that the four helix destabilizing residues Thr 57, Thr 60, Gly 63, and Thr 67 on α-helix H3 of TGF-β3 are responsible for disrupting the formation of this structure in solution.[12] Although no evidence for any loss of helical content was observed in these simulations, it is precisely this area of the protein where mutation of the side chains to hydrogen had the most stabilizing effect upon the induced open conformations of TGF-β1. The localized instability experienced by H3 in adopting the open conformation clearly needs to be overcome by some means. In this study we see how for TGF-β1, this is achieved by reverting to the closed state, while for β3, the helix still remains intact, yet is simply shifted sideways to accommodate the opposite monomer.

The reason TGF-β3 alone is able to tolerate the change in orientation of α-helix H3 lies (at least in part) in the nature of the residues that comprise the preceding loop. In the available experimental structures, position 52 appears semiburied, an environment that clearly suits the highly lipophilic tryptophan residue in TGF-β1. However, during the process of adopting the open conformation and in an attempt to retain native contacts at the dimer interface, this loop becomes more exposed to solvent, and so the environment becomes far better suited to the more polar arginine residue in TGF-β3. Indeed, from a lipophilicity perspective, mutation at this position could not be more dramatic. It is, therefore, not surprising that there are significant differences in behavior, not only between one isoform and another, but also between observations made in solid crystals and those in a polar solvent.

It is likely that the simulations did not sample an adequate amount of time for full helical decomposition to be observed, and that ultimately this is what we would expect to happen. Nevertheless, the findings are still interesting in their own right. We have observed that in solution, TGF-β3 has the ability to adopt a high-energy intermediate conformation in which the two monomers are rotated but α-helix H3 remains intact. TGF-β1 on the other hand, lacks the ability to stabilize such a conformation, and immediately reverts back to the closed state. This implies that kinetics is likely to play a

key role in the transition between open and closed forms of TGF-β isoforms.

Conclusion

We have successfully performed four TMD simulations on two TGF-β isoforms, -β1 and -β3. Both the experimental structures of each isoform and the two homology models remained stable throughout the course of a 11 ns MD simulation in explicit solvent. Furthermore, with the aid of a rmsd weighted restraint, we were able to promote the transition of both isoforms from the closed conformation to one we believe to be more representative of TGF-β3's structure in solution.

Following the simulations from this point, we observed distinct, sequence determined preferences for either the open or the closed form of TGF-β. Transition between the open and closed form is key to the biological function of TGF-β3 because it must adopt the closed conformation to promote assembly of the TGF-β:TβR-I:TβR-II signaling complex as has now been observed.[15] Evidence from NMR, CD, and now TMD simulations indicate that a second, open conformation is preferable in the case of TGF-β3. This is in contrast to the evidence surrounding TGF-β1, which is believed to exist entirely in the closed conformation. Here, we have presented an argument as to why this is the case, and highlighted key mutations at positions 52−58 inclusive that we believe are responsible for the change in behavior.

The protocol outlined in this paper has been shown to distinguish between the conformational preferences of TGF-β1 and -β3, taking just the experimental structure of one isoform as a starting point. Of significant interest now is likely to be the identification of a small subset of point mutations capable of reversing the behavior of a given isoform. This is clearly a potential avenue for further exploration, ideally with the support of experimental evidence to validate any predictions.

Acknowledgment. The authors would like to thank Breast Cancer Campaign for funding this research. D.J.W. would also like to thank Dr. Charles Laughton of the University of Nottingham for use of the Linux cluster on which the calculations were performed

References

(1) Massague, J. *Annu. Rev. Cell Biol.* **1990**, *6*, 597–641.

(2) de Caestecker, M. *Cytokine Growth Factor Rev.* **2004**, *15*, 1–11.

(3) Cheifetz, S.; Hernandez, H.; Laiho, M.; ten Dijke, P.; Iwata, K. K.; Massague, J. *J. Biol. Chem.* **1990**, *265*, 20533–20538.

(4) Massague, J.; Andres, J.; Attisano, L.; Cheifetz, S.; Lopez-Casillas, F.; Ohtsuki, M.; Wrana, J. L. *Mol. Reprod. Dev.* **1992**, *32*, 99–104.

(5) Wrana, J. L.; Attisano, L.; Wieser, R.; Ventura, F.; Massague, J. *Nature* **1994**, *370*, 341–7.

(6) Zúniga, J. E.; Groppe, J. C.; Cui, Y.; Hinck, C. S.; Contreras-Shannon, V.; Pakhomova, O. N.; Yang, J.; Tang, Y.; Mendoza, V.; López-Casillas, F. *J. Mol. Biol.* **2005**, *354*, 1052–1068.

(7) Daopin, S.; Piez, K. A.; Ogawa, Y.; Davies, D. R. *Science* **1992**, *257*, 369–373.

(8) Schlunegger, M. P.; Grütter, M. G. *Nature* **1992**, *358*, 430–434.

(9) Mittl, P. R.; Priestle, J. P.; Cox, D. A.; McMaster, G.; Cerletti, N.; Grütter, M. G. *Protein Sci.* **1996**, *5*, 1261.

(10) Hinck, A. P.; Archer, S. J.; Qian, S. W.; Roberts, A. B.; Sporn, M. B.; Weatherbee, J. A.; Tsang, M. L. S.; Lucas, R.; Zhang, B. L.; Wenker, J. *Biochemistry* **1996**, *35*, 8517–8534.

(11) Archer, S. J.; Bax, A.; Roberts, A. B.; Sporn, M. B.; Ogawa, Y.; Piez, K. A.; Weatherbee, J. A.; Tsang, M. L. S.; Lucas, R. *Biochemistry* **1993**, *32*, 1164–1171.

(12) Bocharov, E. V.; Korzhnev, D. M.; Blommers, M. J. J.; Arvinte, T.; Orekhov, V. Y.; Billeter, M.; Arseniev, A. S. *J. Biol. Chem.* **2002**, *277*, 46273–46279.

(13) Pellaud, J.; Schote, U.; Arvinte, T.; Seelig, J. *J. Biol. Chem.* **1999**, *274*, 7699–7704.

(14) Hart, P. J.; Deep, S.; Taylor, A. B.; Shu, Z.; Hinck, C. S.; Hinck, A. P. *Nat. Struct. Biol.* **2002**, *9*, 203–8.

(15) Groppe, J.; Hinck, C. S.; Samavarchi-Tehrani, P.; Zubieta, C.; Schuermann, J. P.; Taylor, A. B.; Schwarz, P. M.; Wrana, J. L.; Hinck, A. P. *Mol. Cell* **2008**, *29*, 157–168.

(16) Lin, S. J.; Lerch, T. F.; Cook, R. W.; Jardetzky, T. S.; Woodruff, T. K. *Reproduction* **2006**, *132*, 179–190.

(17) Keller, S.; Nickel, J.; Zhang, J. L.; Sebald, W.; Mueller, T. D. *Nat. Struct. Mol. Biol.* **2004**, *11*, 481–488.

(18) Schlitter, J.; Engels, M.; Krüger, P.; Jacoby, E.; Wollmer, A. *Mol. Simul.* **1993**, *10*, 291–308.

(19) Apostolakis, J.; Ferrara, P.; Caflisch, A. *J. Chem. Phys.* **1999**, *110*, 2099.

(20) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. *Nucleic Acids Res.* **2000**, *28*, 235–242.

(21) Case, D. A.; Darden, T. A.; Cheatham, T. E., III; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Merz, K. M.; Wang, B.; Pearlman, D. A.; Crowley, M.; Brozell, S.; Tsui, V.; Gohlke, H.; Mongan, J.; Hornak, V.; Cui, G.; Beroza, P.; Schafmeister, C.; Caldwell, J. W.; Ross, W. S.; Kollman, P. A. *AMBER 8*; University of California: San Francisco, CA, 2004.

(22) Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T. *J. Comput. Chem.* **2003**, *24*, 1999–2012.

(23) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926.

(24) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *Journal Comput. Phys.* **1977**, *23*, 327.

(25) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089.

(26) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684.

(27) Tsui, V.; Case, D. A. *Biopolymers (Nucleic Acid Sci.)* **2001**, *56*, 257–291.

(28) Weiser, J.; Shenkin, P. S.; Still, W. C. *J. Comput. Chem.* **1999**, *20*, 217–230.

(29) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. *J. Comput. Chem.* **2004**, *25*, 1605–1612.

CT800483Q