A potential new, stable state of the E-cadherin strand-swapped dimer in solution

Alexandra Schumann-Gillett[1,2], Alan E. Mark[1,3], Evelyne Deplazes[1,3,*], Megan L. O'Mara[1,2,*]

[1]School of Chemistry and Molecular Biosciences (SCMB), University of Queensland, Brisbane, QLD 4072, Australia

[2]Research School of Chemistry (RSC), The Australian National University, Canberra, ACT 2061, Australia

[3]The Institute for Molecular Biosciences (IMB), University of Queensland, Brisbane, QLD 4072, Australia

* Corresponding authors at: School of Biomedical Sciences, Curtin University, Bentley, WA 6102, Australia; Research School of Chemistry (RSC), The Australian National University, Canberra, ACT 2061, Australia.

E-mail address: evelyne.deplazes@curtin.edu.au (E. Deplazes); megan.o'mara@anu.edu.au (M. L. O'Mara).

Telephone: +61 8 9266 5685 (E. Deplazes); +61 2 6125 3739 (M. L. O'Mara)

Acknowledgements

This work was supported by grants from the Australian Research Council (ARC) to AEM and MLO (DP130102153), and the Medical Advances Without Animals Trust (MAWA) to MLO, ED and ASG. ED is a NHMRC Early Career Research Fellow. This research was undertaken with the assistance of resources provided at the National Computational Infrastructure National Facility systems, housed at the Australian National University, through the National Computational Merit Allocation Scheme supported by the Australian Government.

Abstract

E-cadherin is a transmembrane glycoprotein that facilitates inter-cellular adhesion in the epithelium. The ectodomain of the native structure is comprised of five repeated immunoglobulin-like domains. All E-cadherin crystal structures show the protein in one of three alternative conformations: a monomer, a strand-swapped *trans* homodimer and the so-called X-dimer, which is proposed to be a kinetic intermediate to forming the strand-swapped *trans* homodimer. However, previous studies have indicated that even once the *trans* strand-swapped dimer is formed, the complex is highly dynamic and the E-cadherin monomers may reorient relative to each other. Here, molecular dynamics simulations have been used to investigate the stability and conformational flexibility of the human E-cadherin *trans* strand-swapped dimer. In four independent, 100 ns simulations, the dimer moved away from the starting structure and converged to a previously unreported structure, which we call the Y-dimer. The Y-dimer was present for over 90 % of the combined simulation time, suggesting that it represents a stable conformation of the E-cadherin dimer in solution. The Y-conformation is stabilised by interactions present in both the *trans* strand-swapped dimer and X-dimer crystal structures, as well as additional interactions not found in any E-cadherin dimer crystal structures. The Y-dimer represents a previously unreported, stable conformation of the human E-cadherin *trans* strand-swapped dimer and suggests that the available crystal structures do not fully capture the conformations that the human E-cadherin *trans* homodimer adopts in solution.

Keywords

E-cadherin, Homodimer, Molecular dynamics, Converged structure

Abbreviations

DEER	Double electron-electron resonance
EC	Extracellular cadherin
GROMACS	GROningen MAchine for Chemical Simulation
MD	Molecular dynamics
PDB	Protein Data Bank
RMSD	Root-mean-square deviation
SPC	Simple point charge

Introduction

E-cadherin is a type I classical cadherin that facilitates calcium-dependent adhesion between epithelial cells (Takeichi 1990). This transmembrane glycoprotein is crucial for embryonic development and morphogenesis (Takeichi 1988), as well as suppression of tumour growth and malignant invasion (recently reviewed in (van Roy 2014)). Figure 1 shows the crystal structure of the full mouse E-cadherin ectodomain. It contains two E-cadherin monomers, each consisting of five repeated immunoglobulin-like domains (termed EC domains, numbered 1–5 from N-terminus to C-terminus). Each EC domain is approximately 110 amino acids long. The EC domains are connected by an approximately 10 amino acid-long linker that binds Ca^{2+} *in vivo* (Takeichi 1988; Harrison et al. 2011). The monomers form a strand-swapped *trans* homodimer. Several experimental studies have confirmed that type I classical cadherins form a *trans* homodimer between EC1 domains in solution (Patel et al. 2003; Troyanovsky et al. 2003; Harrison et al. 2005; Zhang et al. 2009). This dimerisation interface has been extensively characterised crystallographically (Shapiro et al. 1995; Boggon et al. 2002; Häussinger et al. 2004; Brasch et al. 2011; Harrison et al. 2011), with structures mainly reported showing a monomer, a strand-swapped dimer or what is suggested to be a kinetic intermediate to forming a strand-swapped dimer, the X-dimer. Table 1 shows fourteen E-cadherin crystal structures that are available in the Protein Data Bank (PDB).

The crystal structures of E-cadherin (and most other type I cadherins) containing the native Trp2 show the six highly conserved N-terminal residues (called the adhesion arm or A*/A-strand) in either an “open” (Häussinger et al. 2004; Parisini et al. 2007;

Harrison et al. 2010; Harrison et al. 2011) or a “closed” conformation (Nagar et al. 1996; Pertz et al. 1999; Harrison et al. 2010). In the open conformation, Trp2 binds within a hydrophobic pocket on the opposing monomer and forms a strand-swapped dimer in the crystal. The hydrophobic pocket is formed by Lys25, Glu89, Asp90 and Met92 (Parisini et al. 2007). In the closed conformation, Trp2 binds into the hydrophobic pocket within its own monomer. In the crystal lattice, the closed monomers form an adhesive dimer mediated by inter-domain contacts, called the X-dimer. (Nagar et al. 1996; Pertz et al. 1999; Harrison et al. 2010). Experimental studies indicate that the X-dimer is a kinetic intermediate in the formation of the mature strand-swapped type I cadherin dimer (Sivasankar et al. 2009; Harrison et al. 2010; Li et al. 2013). Indeed, mutating Trp2 or the residues forming the hydrophobic pocket abolishes strand-swapping (Tamura et al. 1998; Ozawa 2002; Mohamet et al. 2011) and may stabilise the X-dimer conformation (Harrison et al. 2010).

It is widely accepted that type I cadherins share a conserved adhesion mechanism, yet the details remain unclear. This is largely due to the dimerisation pathway involving multiple discrete steps (Dalle Vedove et al. 2015). However, some features of the strand-swapped adhesion mechanism have been inferred from the available crystal structures. Cadherins are translated as inactive protein precursors; extra N-terminal residues are thought to hinder the formation of a salt bridge between Glu89 and Asp1, which prevents Trp2 from docking into its own hydrophobic pocket, or that of another monomer. Removing the N-terminal pro-domain renders the cadherin active and able to bind Trp2. These interactions are primarily stabilised by: a) a salt bridge between Glu89 and Asp1, and b) stacking interactions between Trp2, and Glu89 and Met92. The equilibrium between the closed (intra-) and open (inter-) forms may be affected

by Pro5 and Pro6, which are sterically constrained in the closed conformation (Parisini et al. 2007). Crystallographic structures can only provide snapshots of the protein in specific conformations; limiting the amount of information gained about the multistep dimerisation mechanism. For this reason, molecular dynamics simulations have previously been used to gain insight into the mechanism of E-cadherin dimerisation.

Three studies have used molecular dynamics (MD) simulations of mouse E-cadherin in solution to study the conformational changes and the role of calcium binding in the dimerisation process. Simulations of the monomer and dimer conformations indicate that Ca^{2+} binding promotes N-terminal strand-swapping (Vendome et al. 2011), while other simulations of the monomer confirmed that Ca^{2+} is required to stabilise and rigidify the structure (Cailliez and Lavery 2005). Simulations of a strand-swapped crystal structure suggested that the relative orientation of the monomers was dynamic (Cailliez and Lavery 2006) and indicated that the strand-swapped dimer may adopt a different conformation in solution to that observed in the available crystal structures. However, these simulations were only 12.5 ns in duration and the structures had not converged to a well-defined alternative conformation. In this study, four 100 ns simulations of a crystallographic human EC1–EC2 E-cadherin, modelled in the strand-swapped dimer, have been performed. These simulations converge to reveal a previously unreported conformation of the strand-swapped dimer: the Y-dimer. This conformation has an extensive inter-domain interface that is stabilised by inter-molecular contacts present in both *trans* strand-swapped dimer and X-dimer crystal structures, and additional interactions not present in E-cadherin crystal structures.

Methods

System setup

There is no publically available structure of the full human strand-swapped dimer. However, the Protein Data Bank (PDB) does contain structures of the human EC1–EC2 monomer (PDBid: 2O72) solved to 2.00 Å resolution (Parisini et al. 2007), and the mouse EC1–EC5 E-cadherin strand-swapped dimer (PDBid: 3Q2V) resolved to 3.40 Å (Harrison et al. 2011). The EC1–EC2 domains of mouse and human E-cadherin have 98.1 % sequence similarity and 87.3 % identity, as determined using the Basic Local Alignment Search Tool (Altschul et al. 1990). We constructed a model of the human strand-swapped dimer by superimposing the structure of the human EC1–EC2 domains on the corresponding EC1–EC2 domains of the mouse strand-swapped homodimer. The root mean squared deviation (RMSD) between the human and mouse EC1–EC2 domains was 1.2 Å. To avoid incorporating inappropriate charges in the truncated EC domains, the C-termini were capped with NH_2. The coordinates of the crystallographic Mn^{2+} and Ca^{2+} present in the mouse structure were retained in the human strand-swapped dimer model. While classical force fields reproduce parameters such as the overall ionic charge and ionic radius, they cannot differentiate between coordination states or factors influenced by the quantum mechanical properties of each ion. Furthermore, like many biomolecular force fields, the GROMOS 54A7 force field does not contain specific parameters for Mn^{2+}. Due to these limitations, the parameters for Mg^{2+} were used as an Mn^{2+} mimetic. Mg^{2+} has the same net charge as Mn^{2+} and the ionic radius of the two ions differs by only 0.02 nm. The model EC1–EC2 dimer was placed in the centre of a

rectangular box and periodic boundary conditions were applied so that there was at least 2 nm between the protein and its periodic image. The system was solvated with simple point charge (SPC) water (Berendsen et al. 1981) and neutralised with 527 Na^+ ions and 519 Cl^- ions, to give an overall concentration of 150 mM NaCl.

Molecular dynamics simulation parameters

The simulations were performed using the GROningen MAchine for Chemical Simulation (GROMACS) (van der Spoel et al. 2005) package, version 3.3.3 (Lindahl et al. 2001), and the GROMOS 54A7 force field (Schmid et al. 2011). Each system was simulated under periodic boundary conditions in a rectangular simulation box. The temperature of the system was maintained by coupling the protein, together with the solvent and ions, to an external temperature bath at 300 K with a coupling constant, τ_T = 0.1 ps, using a Berendsen thermostat (Berendsen et al. 1984). The pressure was maintained at 1 bar by weakly coupling the system to an isotropic pressure bath using an isothermal compressibility of 4.5 x 10^{-5} bar^{-1} and a coupling constant, τ_P = 0.5 ps. During the simulations, the length of all covalent bonds were constrained using the LINCS algorithm (Hess et al. 1997). The SETTLE algorithm (Miyamoto and Kollman 1992) was used to constrain the geometry of water molecules. To enable a 4 fs time step to be used, the mass of hydrogen atoms was increased to 4 a.m.u. by transferring mass from the atom to which it was attached. This allows the use of a larger time step to integrate the equation of motion without significantly affecting the thermodynamic properties of the system (Feenstra et al. 1999). Non-bonded interactions were calculated using a twin-range cut-off. Interactions within the short-range cut-off of 0.8 nm were updated every time step.

Interactions within the long-range cut-off of 1.4 nm were updated every second step, together with the pair list. To correct for the truncation of electrostatic interactions beyond the 1.4 nm long-range cut-off, a reaction field correction was applied using an effective dielectric permittivity value, ε_r = 78.5 (Tironi et al. 1995). The atomic coordinates and energies were saved every 50 ps for analysis. All images were produced using Visual Molecular Dynamics (VMD) (Humphrey et al. 1996).

To initiate the simulations, the system was energy minimised using a steepest descent algorithm and equilibrated over 5 ns. The position restraints were successively released during equilibration. Four independent, unrestrained simulations of 100 ns duration were performed, each starting with a different random velocity. These are referred to as MD1 to MD4. To investigate the influence of the presence of Mg^{2+}, the choice of force field, and the choice of MD engine on the dimer conformation, four additional simulations were performed. The simulation conditions were identical to the original simulations except a) two replicate simulations lacked Mg^{2+}, b) one simulation was performed using GROMACS version 5.1.4, and c) one simulation was performed using GROMACS version 5.1.4 and the Amber99SB-ILDN protein (Lindorff-Larsen et al. 2010) force field.

Analysis

Root-mean-square deviation (RMSD). The backbone atoms of the protein (Cα, C, N and O) were fit to a reference conformation using a least-squares method. The RMSD of these atoms was then calculated for each frame in a trajectory, using the method employed by Kabsch (Kabsch 1976). Unless otherwise stated, the reference

conformation was the first frame of the unrestrained simulations. The average RMSD and standard deviation were calculated for each simulation.

Cluster analysis. The four trajectories were combined into a single data set containing 1600 structures for cluster analysis. The cluster algorithm of Daura *et al.* (Daura et al. 1999a; Daura et al. 1999b) was used to group conformations sampled during the simulations into clusters, using a backbone neighbour RMSD cut-off of 2.5 Å.

Dimer interface contact maps. To define which residues form the dimer interface, the relative proportion of simulation time that each residue on one monomer was within 4 Å of each residue on the opposing monomer was calculated for the human E-cadherin EC1–EC2 crystallographic construct (the starting structure for the simulations) and the final 60 ns of each simulation. The data was normalised against the total number of frames.

Results

Conformations of the human EC1–EC2 E-cadherin strand-swapped dimer in solution

The backbone RMSD of the dimer during each of the four 100 ns simulations of the human EC1–EC2 strand-swapped dimer, with respect to the starting structure, is shown in Fig. 2. Representative snapshots indicating the time evolution of the dimer during simulation MD1 are also shown (Fig. 2). Consistent with Fig. 2, the Mg^{2+} and inter-domain Ca^{2+} ions remained bound throughout MD1–MD4. During all

simulations, the relative conformation of each monomer shifts significantly during the first 30 ns, reflected by the large increase in the RMSD during this period. The backbone RMSD then stabilises at a value between 12.5 and 24 Å over the remaining 70 ns, depending on the simulation. In all replicates, the RMSD stabilises after approximately 40 ns of simulation. The average RMSD after 40 ns for each simulation was 16.9 ± 1.3 Å (MD1), 20.1 ± 1.1 Å (MD2), 17.9 ± 2.9 Å (MD3) and 19.7 ± 1.3 Å (MD4).

We used cluster analysis to confirm that the large backbone RMSDs represented a conformational change between the monomers, and to identify the prevalent conformations of the dimer in solution. A total of 1600 structures from the combined simulations were fit to the strand-swapped starting conformation, and then clustered using a 2.5 Å backbone RMSD cut-off. The most populated cluster contained 90.9 % of the structures from the combined simulations. The central conformation from this cluster is shown in Fig. 3b. It is significantly different than the strand-swapped starting conformation, as shown in Fig. 3a. The middle structure from the most populated cluster represents the most dominant conformation of the human E-cadherin dimer in our simulations. Because of its characteristic shape, we refer to this structure as the Y-dimer. This is the first time that the Y-dimer conformation of the E-cadherin strand-swapped dimer has been reported. The second- and third-most prevalent conformations were present for 1.7 % and 1.3 % of the combined trajectory respectively (see Online Resource Fig. 1).

To confirm that the dimer converges to the Y-dimer conformation in all simulations, we calculated the backbone RMSD between the starting structure and central

conformation of the most populated cluster. Figure 4 shows the RMSD plotted over the duration of each simulation. As expected, the plot follows an inverted trend compared to the RMSD calculated with respect to the starting strand-swapped conformation (Fig. 2). The RMSD in Fig. 4 is relatively high during the first 20–30 ns, reflecting the reorientation of the monomers as they form the Y-dimer conformation. The spike down to 0 Å RMSD for MD2 at approximately 22 ns corresponds to the reference structure that was used to perform the calculation. In each independent simulation, the RMSD stabilises after approximately 30 ns, as the Y-dimer conformation is adopted. To illustrate the relatively small difference in the converged conformation across each simulation, the backbone RMSD for the last frame of each simulation, calculated with respect to the Y-dimer, was 4.8 Å (MD1), 4.3 Å (MD2), 5.6 Å (MD3) and 3.2 Å (MD4). The standard deviation between these structures is 1.0 Å. Taken together, the data indicate that the conformation of the human E-cadherin EC1–EC2 strand-swapped dimer has converged to the Y-dimer conformation in all simulations.

Characterising the Y-dimer interface

The Y-dimer, shown in Fig. 3b, is the most prevalent conformation adopted by the E-cadherin strand-swapped dimer in all of our simulations. It is clear from Fig. 3 that there is a larger contact interface between the EC1 domains of each monomer in the Y-dimer than there is in the starting structure (Fig. 3a), or any of the other conformations identified here (see Online Resource Fig. 1c–d). Figure 5a gives a 2-dimensional contact map of the residues involved in the dimer interface. It shows the frequency that each residue on one monomer is within 4 Å of each residue on the

opposing monomer for the starting structure (the human crystallographic construct) (Fig. 5a). In Fig. 5b, the contact map is calculated over the final 60 ns of MD1. The contact maps for all four independent simulations are remarkably similar (Online Resource Fig. 2b–e).

The contact maps are largely symmetric along the x = y diagonal, reflecting the paired interactions that take place upon dimerisation. For example, Trp2 of one monomer binds in the hydrophobic pocket of the opposing monomer, and vice versa. The inter-molecular contact map shown in Fig. 5a reflects the N-terminal inter-domain interactions characterised crystallographically in the *trans* strand-swapped dimer. These are isolated to EC1 and include interactions between Trp2 and residues around the hydrophobic pocket (Lys25, Glu89, Asp90 and Met92), namely those between Trp2 and Lys25, Ser78, His79, Ala80, Glu89, Asp90, Pro91, Met92 and Ile94. The contact map also captures interactions between the opposing contact residue pairs: Asp1—Asn27; Val3—Lys25; Trp2—Ser37; and Pro5—Ile24, and their neighbouring residues.

The inter-molecular interactions seen in the initial strand-swapped conformation are retained and expanded when the human E-cadherin dimer shifts from this conformation to the Y- conformation during the simulations. There are seven red–orange regions in Fig. 5b (and Online Resource Fig. 2c–e), which correspond to frequent (above 50 %) inter-molecular interactions in the Y-dimer. These regions lie along the x = y diagonal for residues 1–10; around the residues involved in inter-molecular Trp2 binding (crystallographically characterised as Trp2, Lys25, Glu89, Asp90 and Met92); and around the opposing contact residue pair 19—140.

The frequent interactions shown in the contact maps reflect those that maintain the Y-conformation of the dimer in each simulation. The interactions seen along the x = y diagonal for residues 1–10, form the adhesion arm of each monomer and are dominated by hydrophobic interactions between Trp2, Val3, Ile4, Pro5, Pro6 and Ile7 (see Online Resource Fig. 3). In our simulations, Trp2 binds to the hydrophobic pocket of the opposing monomer. This, together with recurrent inter-molecular hydrogen bonds between Asp1—Asn27 and Val3—Lys25, anchors the N-terminus and serve as a pivot point during dimer conformational change from the starting structure to the Y-dimer. Hydrogen bond formation between Ser8 of the opposing monomers stabilises the C-terminal end of the adhesion arm. This interaction is formed while the dimer is shifting from the starting strand-swapped conformation to the Y-dimer, and remains stable throughout the simulations. Figure 5B also shows contacts between Lys19 and Asn140, which can be attributed to hydrogen bonding between the side chain of Lys19 and proton-accepting groups on Asn140 (see Online Resource Fig. 4). Asn140 is located on the EC2 domain near the calcium-binding site. These inter-molecular interactions near the calcium-binding site are reminiscent of the interactions between E-cadherin monomers in the X-dimer conformation.

For comparison, we used a mouse E-cadherin crystal structure (PDBid 3LNG) (Harrison et al. 2010) to generate a contact map of the X-dimer interactions Online Resource Fig. 5). The X-dimer is structurally distinct from the strand-swapped dimer: a key difference is that Trp2 does not dock into hydrophobic pocket of the other monomer. This is demonstrated by the small degree of overlap between its contact map and that of either the starting structure or the Y-dimer. One common feature of

both the Y-dimer and X-dimer contact maps are the interactions between and around residue pair 19—140 (see Online Resource Fig. 2b–f). As Trp2 does not bind to the opposing monomer in the X-dimer (there are no contacts between Trp2 and the hydrophobic pocket residues Lys25, Glu89, Asp90 and Met92 in Online Resource Fig. 2f), the X-dimer contact map shows only two interactions along the adhesion arm: between residue pair 5—4 and 5—21. All other interactions in the X-dimer contact map are between residues near the calcium-binding sites. When taken together, these findings reinforce the conclusion that our simulations have converged on an unreported conformation of the E-cadherin *trans* dimer: the Y-dimer.

Discussion

Previous simulations have suggested that the E-cadherin *trans* strand-swapped dimer adopts different conformations in crystal structures and solution (Cailliez and Lavery 2006). We performed four independent MD simulations, 100 ns each, to investigate the conformations adopted by a human E-cadherin crystallographic *trans* strand-swapped dimer construct in solution. RMSD calculations, cluster analysis and inter-molecular contact maps indicate that the simulations have converged to a single structural conformation of the E-cadherin *trans* strand-swapped dimer that has not been reported previously, referred to as the Y-dimer. In this conformation, the characteristic, biologically-relevant strand-swapping (Trp2 exchange) exhibited by mature E-cadherin *trans* dimers is retained. The Y-dimer also has extensive inter-molecular contacts along the length of EC1. Only one of these contacts (between Lys19 and Asn140) is shared between the Y-dimer and the other biologically-relevant crystal structure of the E-cadherin dimer, the X-dimer. This interaction, and other

interactions that exclusively define the Y-dimer interface, have not been studied experimentally (for example, in mutational studies).

In MD simulations performed by Cailliez and Lavery, the structure of a mouse E-cadherin strand-swapped dimer diverged from the starting crystal structure (Cailliez and Lavery 2006). Cailliez and Lavery reported that the EC1 domains of each monomer came together over the 12.5 ns simulation. The average structure over the final 2.5 ns of this simulation appears to represent a conformation intermediate between our initial human strand-swapped starting conformation and the Y-dimer characterised here. To investigate whether the Y-dimer conformation observed in MD1–MD4 was dependent on the choice of MD engine or force field, two additional simulations were performed. Here, the simulation conditions were identical to MD1–MD4, except one simulation used GROMACS 5.1.4 in conjunction with GROMOS 54a7, while the other used GROMACS 5.1.4 and the Amber99SB-ILDN protein (Lindorff-Larsen et al. 2010) force field. As shown in Online Resource Fig. 6, the Y-dimer was formed in both of these simulations.

It is important to note that our starting structure contained Mg^{2+} ions as a mimetic for the Mn^{2+} ions present in the EC1–EC5 mouse E-cadherin crystal structure. An Mn^{2+} mimetic was not present in the simulations performed by Cailliez and Lavery. To investigate the effect of these ions on the dimer conformation we observed in our simulations, we performed two additional simulations in which the Mg^{2+} ions were removed. After 30 ns of simulation the Mg^{2+}-free dimer diverged from the strand-swapped starting conformation and moved towards the Y-dimer conformation (see Online Resource Fig. 7). The backbone RMSD of the Mg^{2+}-free dimer was

comparable to that observed with bound Mg^{2+}. Furthermore, Trp2 strand-swapping was retained. These simulations of the E-cadherin strand-swapped dimer in solution show that the structure converges to the Y-dimer irrespective of the presence of a divalent cation in the position observed in the full mouse E-cadherin crystal structure. This work suggests that the E-cadherin *trans* strand-swapped dimer conformation captured crystallographically may differ from the conformation in solution.

The Y-dimer interface displayed unique, persistent hydrophobic contacts along the EC1 adhesion arm—between Trp2, Val3, Ile4, Pro5, Pro6 and Ile7 of each monomer (see Fig. 5, Online Resource Fig. 2). Mutational studies of residues unique to the Y-dimer interface could further validate the presence of the Y-dimer experimentally. For instance, we anticipate that mutating the hydrophobic residues along the adhesion arm (Trp2, Val3, Ile4, Pro5, Pro6 and Ile7) to hydrophilic residues would disrupt the Y-dimer interface, altering any structural results obtained by methods such as nuclear magnetic resonance spectroscopy. Work by Vendome (Vendome et al. 2014) using double electron-electron resonance (DEER) spectroscopy reports a single distance measurement of 7 nm between the nitroxide-labelled 135C residue for wild-type E-cadherin. This was the upper limit of accurate peak width determination, as stated in the paper. The predicted distance based on a wild-type strand-swapped E-cadherin crystal structure was approximately 6.3 nm. The approximate distance between unlabelled residue 135 of each monomer in the Y-dimer conformation (data not shown) is 4 nm. In Vendome and co-workers' paper, E-cadherin monomers were deemed "invisible" in the DEER experiment, and the mobility and orientation of the spin labels is not fully described. However, it is interesting to note that the crystal structure of a mouse (neural) N-cadherin amino acid-insertion mutant solved in

Vendome and co-workers' study resembles the Y-dimer conformation seen in the simulations performed here. The predicted crystallographic and DEER-measured distance between nitroxide-labelled residues 135 for the N-cadherin amino acid-insertion mutant was approximately 4 nm (Vendome et al. 2014); the same as the unlabelled distance in the Y-dimer. Furthermore, it has been proposed that structures of human and mouse E-cadherin captured in crystals could reflect local minima on a relatively shallow free energy surface (Parisini et al. 2007). Our simulations may sample a different region of the free energy landscape, adding to the discrepancy between the mouse crystal and DEER structures, and the human Y-conformation seen in this study. Overall, this suggests that further work is needed to fully understand the conformation that the human E-cadherin *trans* strand-swapped dimer adopts in solution, and under physiological conditions.

References

Altschul SF, Gish W, Miller W, et al (1990) Basic Local Alignment Search Tool. J Mol Biol 215:403–410. doi: 10.1006/jmbi.1990.9999

Berendsen HJC, Postma JPM, van Gunsteren WF, et al (1984) Molecular dynamics with coupling to an external bath. J Chem Phys 81:3684. doi: 10.1063/1.448118

Berendsen HJC, Postma JPM, van Gunsteren WF, Hermans J (1981) Interaction models for water in relation to protein hydration. In: Pullman B (ed) Intermolecular forces. Springer Netherlands, pp 331–342

Boggon TJ, Murray J, Chappuis-Flament S, et al (2002) C-cadherin ectodomain structure and implications for cell adhesion mechanisms. Science (80-) 296:1308–1313. doi: 10.1126/science.1071559

Brasch J, Harrison OJ, Ahlsen G, et al (2011) Structure and binding mechanism of vascular endothelial cadherin: a divergent classical cadherin. J Mol Biol 408:57–73. doi: 10.1016/j.jmb.2011.01.031

Cailliez F, Lavery R (2005) Cadherin mechanics and complexation: the importance of calcium binding. Biophys J 89:3895–3903. doi: 10.1529/biophysj.105.067322

Cailliez F, Lavery R (2006) Dynamics and stability of E-cadherin dimers. Biophys J 91:3964–71. doi: 10.1529/biophysj.106.087213

Dalle Vedove A, Lucarelli AP, Nardone V, et al (2015) The X-ray structure of human P-cadherin EC1-EC2 in a closed conformation provides insight into the type I cadherin dimerization pathway. Acta Crystallogr Sect F Struct Biol Commun 71:371–380. doi: 10.1107/S2053230x15003878

Daura X, Gademann K, Jaun B, et al (1999a) Peptide folding: when simulation meets experiment. Angew Chem Int Ed 38:236–240. doi: 10.1002/(Sici)1521-

3773(19990115)38:1/2<236::Aid-Anie236>3.0.Co;2-M

Daura X, van Gunsteren WF, Mark AE (1999b) Folding-unfolding thermodynamics of a β-heptapeptide from equilibrium simulations. Proteins 34:269–280. doi: 10.1002/(SICI)1097-0134(19990215)34:3

Feenstra KA, Hess B, Berendsen HJC (1999) Improving efficiency of large time-scale molecular dynamics simulations of hydrogen-rich systems. J Comput Chem 20:786–798. doi: 10.1002/(SICI)1096-987X(199906)20:8<786::AID-JCC5>3.0.CO;2-B

Harrison OJ, Bahna F, Katsamba PS, et al (2010) Two-step adhesive binding by classical cadherins. Nat Struct Mol Biol 17:348–357. doi: 10.1038/nsmb.1784

Harrison OJ, Corps EM, Kilshaw PJ (2005) Cadherin adhesion depends on a salt bridge at the N-terminus. J Cell Sci 118:4123–4130. doi: 10.1242/jcs.02539

Harrison OJ, Jin X, Hong S, et al (2011) The extracellular architecture of adherens junctions revealed by crystal structures of type I cadherins. Structure 19:244–256. doi: 10.1016/j.str.2010.11.016

Häussinger D, Ahrens T, Aberle T, et al (2004) Proteolytic E-cadherin activation followed by solution NMR and X-ray crystallography. Embo J 23:1699–1708. doi: 10.1038/sj.emboj.7600192

Hess B, Bekker H, Berendsen HJC, Fraaije JGEM (1997) LINCS: A linear constraint solver for molecular simulations. J Comput Chem 18:1463–1472. doi: 10.1002/(Sici)1096-987x(199709)18:12<1463::Aid-Jcc4>3.0.Co;2-H

Humphrey W, Dalke A, Schulten K (1996) VMD: Visual molecular dynamics. J Mol Graph 14:27-28-38. doi: 10.1016/0263-7855(96)00018-5

Kabsch W (1976) A solution for the best rotation to relate two sets of vectors. Acta Crystallogr Sect A 32:922–923. doi: 10.1107/S0567739476001873

Li Y, Altorelli NL, Bahna F, et al (2013) Mechanism of E-cadherin dimerization probed by NMR relaxation dispersion. Proc Natl Acad Sci U S A 110:16462–16467. doi: 10.1073/pnas.1314303110

Lindahl E, Hess B, van der Spoel D (2001) GROMACS 3.0: A package for molecular simulation and trajectory analysis. J Mol Model 7:306–317. doi: 10.1007/s008940100045

Lindorff-Larsen K, Piana S, Palmo K, et al (2010) Improved side-chain torsion potentials for the Amber ff99SB protein force field. Proteins 78:1950–8. doi: 10.1002/prot.22711

Miyamoto S, Kollman PA (1992) Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. J Comput Chem 13:952–962. doi: 10.1002/jcc.540130805

Mohamet L, Hawkins K, Ward CM (2011) Loss of function of E-cadherin in embryonic stem cells and the relevance to models of tumorigenesis. J Oncol 2011:352616. doi: 10.1155/2011/352616

Nagar B, Overduin M, Ikura M, Rini JM (1996) Structural basis of calcium-induced E-cadherin rigidification and dimerization. Nature 380:360–364. doi: 10.1038/380360a0

Ozawa M (2002) Lateral dimerization of the E-cadherin extracellular domain is necessary but not sufficient for adhesive activity. J Biol Chem 277:19600–19608. doi: 10.1074/jbc.M202029200

Parisini E, Higgins JM, Liu JH, et al (2007) The crystal structure of human E-cadherin domains 1 and 2, and comparison with other cadherins in the context of adhesion mechanism. J Mol Biol 373:401–411. doi: 10.1016/j.jmb.2007.08.011

Patel SD, Chen CP, Bahna F, et al (2003) Cadherin-mediated cell-cell adhesion:

sticking together as a family. Curr Opin Struct Biol 13:690–698. doi: 10.1016/j.sbi.2003.10.007

Pertz O, Bozic D, Koch AW, et al (1999) A new crystal structure, Ca^{2+} dependence and mutational analysis reveal molecular details of E-cadherin homoassociation. Embo J 18:1738–1747. doi: 10.1093/emboj/18.7.1738

Schmid N, Eichenberger AP, Choutko A, et al (2011) Definition and testing of the GROMOS force-field versions 54A7 and 54B7. Eur Biophys J with Biophys Lett 40:843–856. doi: 10.1007/s00249-011-0700-9

Shapiro L, Fannon AM, Kwong PD, et al (1995) Structural basis of cell-cell adhesion by cadherins. Nature 374:327–337. doi: 10.1038/374327a0

Sivasankar S, Zhang Y, Nelson WJ, Chu S (2009) Characterizing the initial encounter complex in cadherin adhesion. Structure 17:1075–1081. doi: 10.1016/j.str.2009.06.012

Takeichi M (1990) Cadherins: a molecular family important in selective cell-cell adhesion. Annu Rev Biochem 59:237–52. doi: 10.1146/annurev.bi.59.070190.001321

Takeichi M (1988) The cadherins: cell-cell adhesion molecules controlling animal morphogenesis. Development 102:639–655.

Tamura K, Shan WS, Hendrickson WA, et al (1998) Structure-function analysis of cell adhesion by neural (N-) cadherin. Neuron 20:1153–1163. doi: 10.1016/S0896-6273(00)80496-1

Tironi IG, Sperb R, Smith PE, van Gunsteren WF (1995) A generalized reaction field method for molecular dynamics simulations. J Chem Phys 102:5451–5459. doi: 10.1063/1.469273

Troyanovsky RB, Sokolov E, Troyanovsky SM (2003) Adhesive and lateral E-

cadherin dimers are mediated by the same interface. Mol Cell Biol 23:7965–7972. doi: 10.1128/MCB.23.22.7965-7972.2003

van der Spoel D, Lindahl E, Hess B, et al (2005) GROMACS: Fast, flexible, and free. J Comput Chem 26:1701–1718. doi: 10.1002/jcc.20291

van Roy F (2014) Beyond E-cadherin: roles of other cadherin superfamily members in cancer. Nat Rev Cancer 14:121–34. doi: 10.1038/nrc3647

Vendome J, Felsovalyi K, Song H, et al (2014) Structural and energetic determinants of adhesive binding specificity in type I cadherins. Proc Natl Acad Sci U S A 111:E4175-84. doi: 10.1073/pnas.1416737111

Vendome J, Posy S, Jin X, et al (2011) Molecular design principles underlying β-strand swapping in the adhesive dimerization of cadherins. Nat Struct Mol Biol 18:693–700. doi: 10.1038/nsmb.2051

Zhang YX, Sivasankar S, Nelson WJ, Chu S (2009) Resolving cadherin interactions and binding cooperativity at the single-molecule level. Proc Natl Acad Sci U S A 106:109–114. doi: 10.1073/pnas.0811350106

Tables

PDB code	Asymmetric unit	Sequence alterations	Organism, Reference
1Q1P	Open monomer	r2–213	*Mus musculus*, (Häussinger et al. 2004)
2O72	Open monomer	r1–213	*Homo sapiens*, (Parisini et al. 2007)
2QVF	Open monomer	r1–213	*Mus musculus*, (to be published)
3LNE	Open monomer	r1–213; K14E	*Mus musculus*, (Harrison et al. 2010)
3Q2L	Strand-swapped dimer	r1–213; V81D	*Mus musculus*, (Harrison et al. 2011)
3Q2V	Strand-swapped dimer	r1–536	*Mus musculus*, (Harrison *et al.*, 2011)
3Q2N	Strand-swapped dimer	r1–213; L175D	*Mus musculus*, (Harrison *et al.*, 2011)
1EDH	X-dimer	r3–213	*Mus musculus*, (Nagar et al. 1996)
1FF5	X-dimer	r(-1)–218	*Mus musculus*, (Pertz et al. 1999)
3LNF	X-dimer	r3–213; W2A, K14E	*Mus musculus*, (Harrison et al. 2010)

3LNG	X-dimer	r2–213	*Mus musculus*, (Harrison et al. 2010)
3LNH	X-dimer	r5–213; W2A	*Mus musculus*, (Harrison et al. 2010)
3LNI	X-dimer	r1–213; E89A	*Mus musculus*, (Harrison et al. 2010)
3QRB	X-dimer	r1–213; P5A, P6A	*Mus musculus*, (Vendome et al. 2011)

Table 1 Details of fourteen E-cadherin crystal structures available in the PDB

Figures

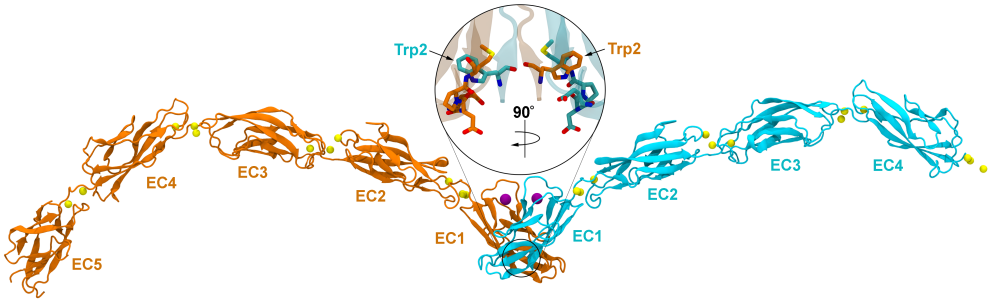

Fig. 1 The crystal structure of the full mouse E-cadherin ectodomain (PDBid: 3Q2V)(Harrison et al. 2011) showing the resolved residues. Each monomer consists of five extracellular cadherin (EC) domains. Trp2 from each of the N-terminal EC domains binds to the acceptor pocket in the opposing domain to form a strand-swapped dimer (see inset). Trp2 and acceptor pocket residues are shown in stick representation. Ligand calcium and manganese ions are shown as yellow and purple spheres, respectively. E-cadherin monomers are in orange or blue cartoon representation

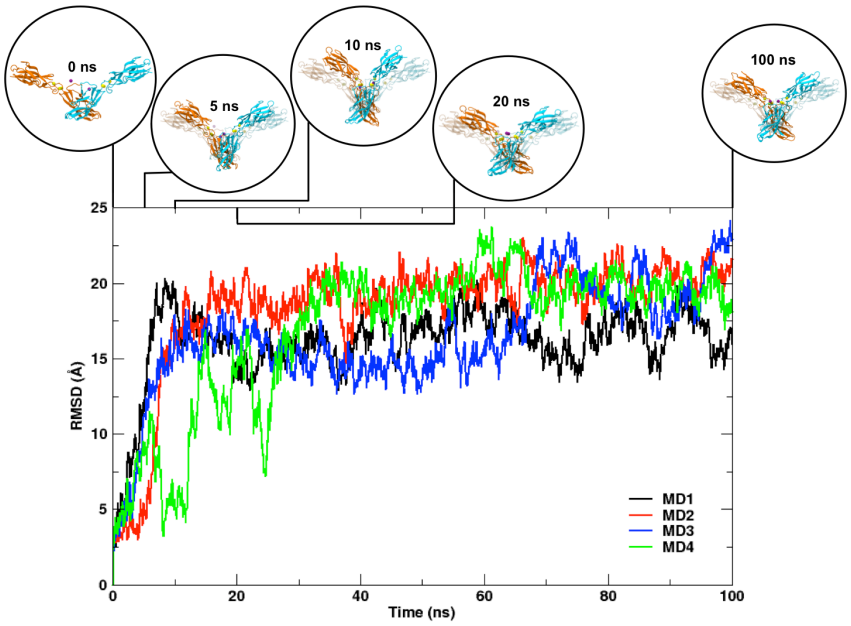

Fig. 2 Backbone RMSD of the E-cadherin strand-swapped dimer during each of the four 100 ns simulations, with snapshots showing the time evolution of the dimer conformation during the first simulation (MD1). The RMSD was calculated with respect to the crystallographic human strand-swapped dimer construct. The snapshots show the initial (0 ns) structure in a transparent representation for reference, and the dimer at a given time point in the simulation in opaque representation. Each monomer is shown in orange or blue cartoon representation, with Mg^{2+} shown as purple spheres and Ca^{2+} as yellow spheres. Solvent molecules were excluded from the figure for clarity

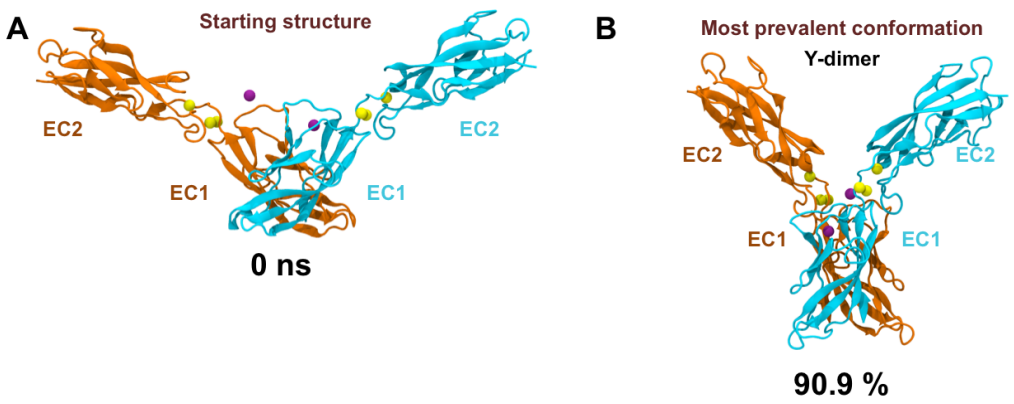

Fig. 3 The starting conformation (a) and the central conformation from the most prevalent cluster (b) of the E-cadherin strand-swapped dimer for the combined simulations determined using cluster analysis with a 2.5 Å RMSD cut-off. We refer to the structure shown in B as the Y-dimer

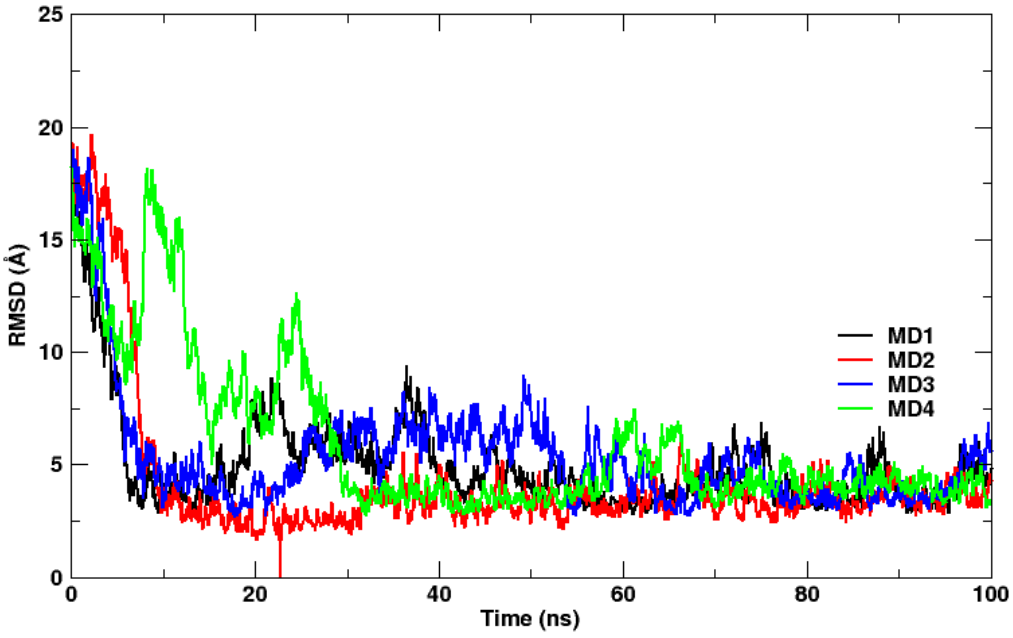

Fig. 4 Backbone RMSD of the E-cadherin strand-swapped dimer for each simulation, calculated with respect to the Y-dimer

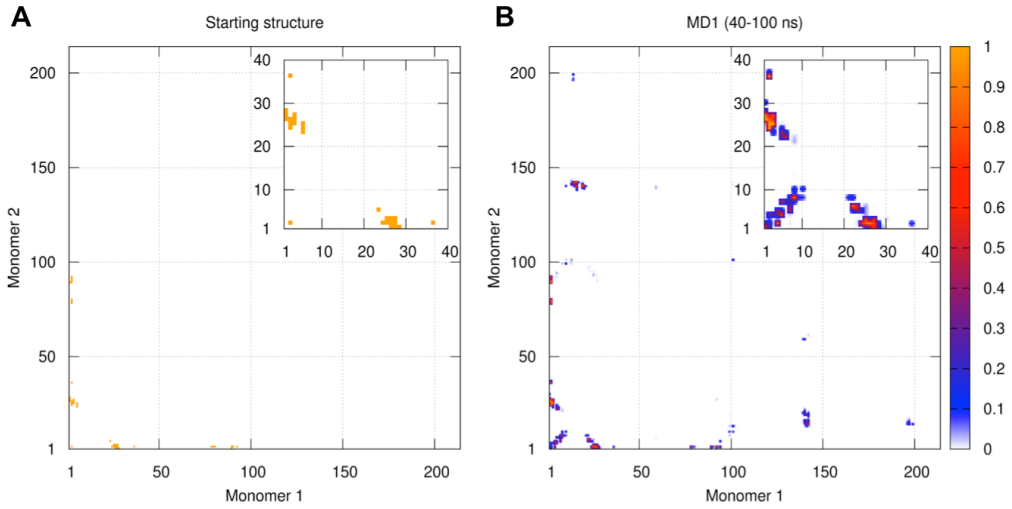

Fig. 5 Contact maps for the dimerisation interface of (a) the starting structure, which was the crystallographic human EC–EC2 strand-swapped dimer construct, and (b) the dimer from the last 60 ns of MD1. The contact map was calculated as the normalised frequency for which each residue of one monomer is within 4 Å of each residue of the opposing monomer. The data was calculated across all residues of both monomers. The data is normalised against the total number of frames in the simulation. The inset in each plot shows an enlarged version of the contact frequency data for residues 1–40. The insets do not hide any data points on the larger maps

A potential new, stable state of the E-cadherin strand-swapped dimer in solution

Supporting Information

Alexandra Schumann-Gillett[1,2], Alan E. Mark[1,3], Evelyne Deplazes[1,3,*], Megan L. O'Mara[1,2,*]

[1]School of Chemistry and Molecular Biosciences (SCMB), University of Queensland, Brisbane, QLD 4072, Australia

[2]Research School of Chemistry (RSC), The Australian National University, Canberra, ACT 2061, Australia

[3]The Institute for Molecular Biosciences (IMB), University of Queensland, Brisbane, QLD 4072, Australia

* Corresponding authors
E-mail address: evelyne.deplazes@curtin.edu.au (E. Deplazes);
megan.o'mara@anu.edu.au (M. L. O'Mara).

Telephone: +61 8 9266 5685 (E. Deplazes); +61 2 6125 3739 (M. L. O'Mara)

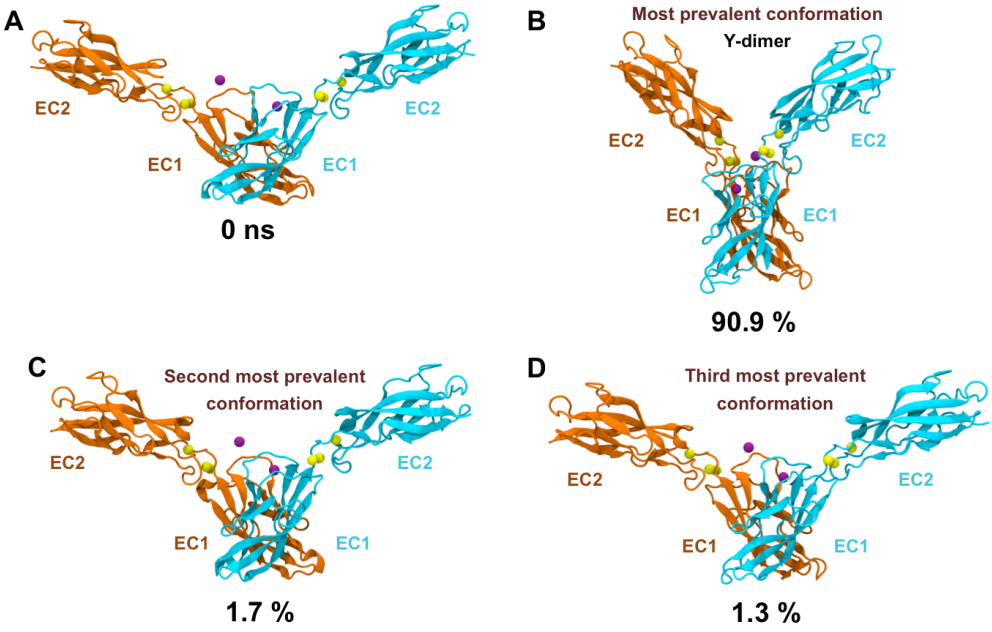

Figure 1: The starting (a) and most prevalent (b–d) conformations of the E-cadherin strand-swapped dimer for the combined simulations, determined using cluster analysis with a clustering cut-off of 2.5 Å (backbone RMSD) on 1600 structures. The structure shown in b corresponds to the Y-dimer. Each monomer is shown as either blue or orange ribbons. Ca^{2+} and Mg^{2+} ions are shown as yellow and magenta spheres, respectively.

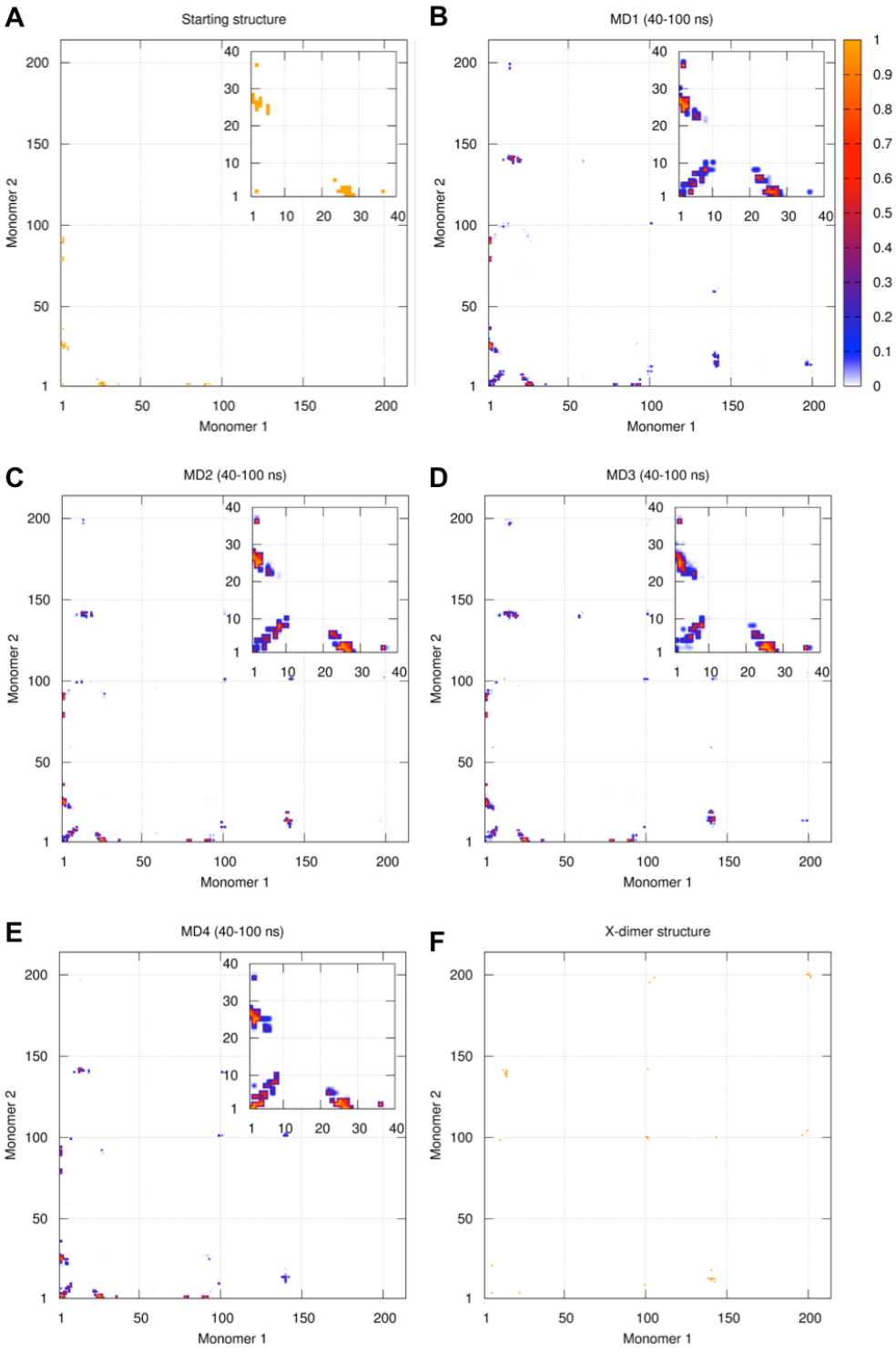

Figure 2: Dimerisation interface contact maps for (a) the starting structure, (b–e) the structure from 40–100 ns of simulation MD1 to MD4, and (f) a murine E-cadherin X-dimer crystal structure (PDB code 3LNG). The data are shown as the frequency that

each residue on one monomer is within 4 Å of each residue on the opposing monomer in the dimer. For plots B–E, the data is normalised against the total number of frames, giving a relative frequency of 0–1 for each data point. The inset in each plot shows an enlarged version of the contact frequency data for residues 1–40. The insets do not overlap with any data points.

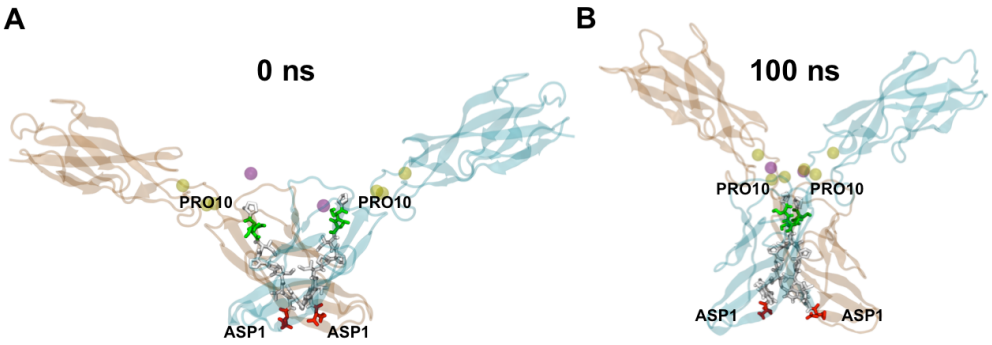

Figure 3: Representative structures of the E-cadherin EC1–EC2 strand-swapped dimer conformation at the start (a) and end (b) of the simulations. The A-strand is shown as sticks, coloured by residue type. Green indicates polar residues, white indicates non-polar residues and red indicates acidic residues.

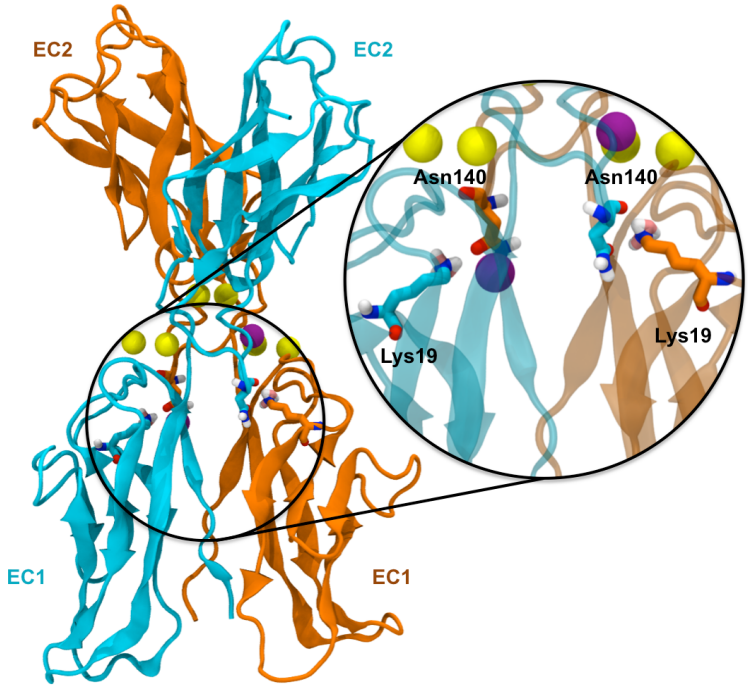

Figure 4: Representative snapshot showing the locations of Lys19 on one monomer and Asn140 on the opposing monomer in the Y-dimer conformation. Lys19 and Asn140 are shown as sticks, coloured by atom type.

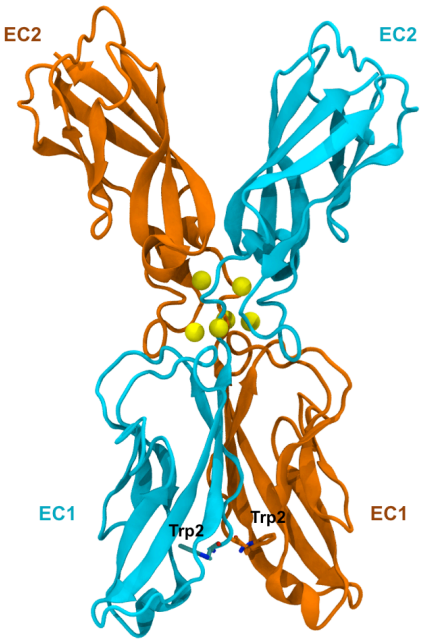

Figure 5: X-dimer crystal structure (PDB code 3LNG) (Harrison et al. 2010) used to generate the X-dimer contact map (Figure S2F). Each monomer is shown as either an orange or a blue ribbon; Ca^{2+} are shown as yellow spheres; and Trp2 from each monomer is shown in stick representation, coloured by atom.

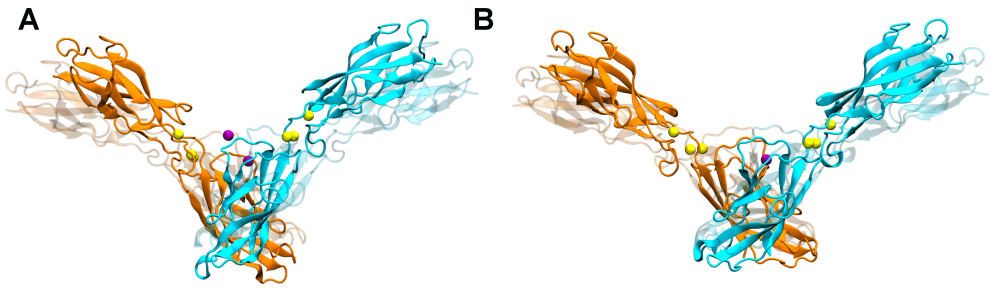

Figure 6: The E-cadherin strand-swapped dimer forms a Y-dimer in simulations using (a) GROMACS 5.1.4 and (b) GROMACS 5.1.4 and the Amber99SB-ILDN force field. The Y-dimer and its associated ions are opaque; as a reference, the starting strand-swapped dimer protein conformation is transparent.

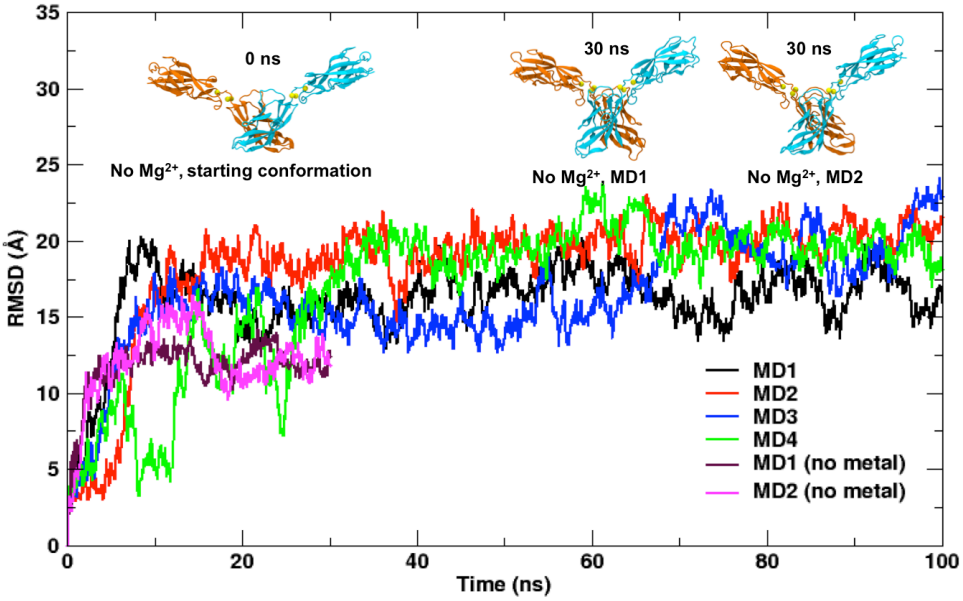

Figure 7: Root-mean-square deviation of the E-cadherin *trans* strand-swapped dimer with and without bound Mg^{2+}. The snapshots show the starting and end conformations of the dimer from the simulations without Mg^{2+}.