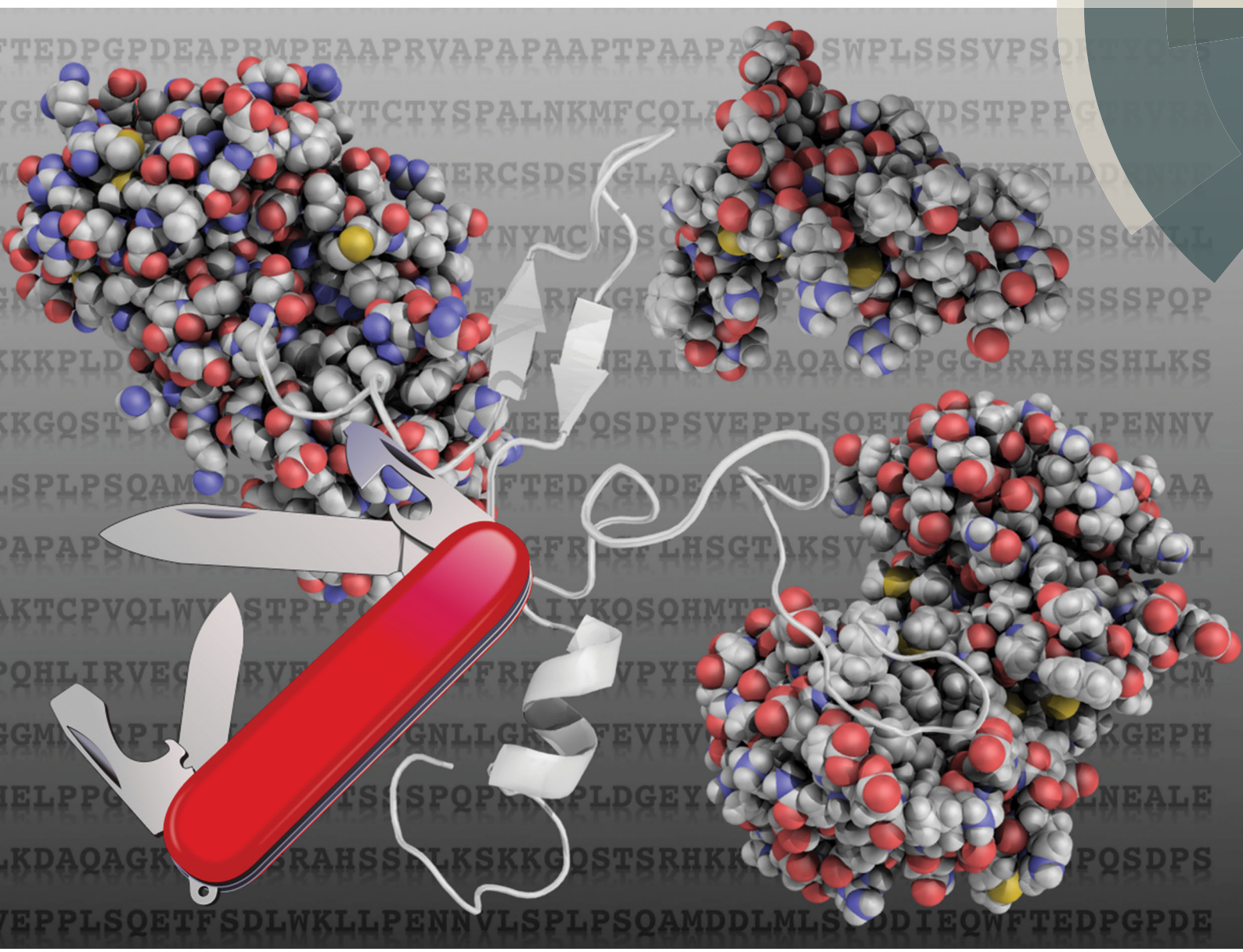
Volume 19 | Number 32 | 28 August 2017 | Pages 21213–21906

PCCP

Physical Chemistry Chemical Physics

rsc.li/pccp



ISSN 1463-9076

ROYAL SOCIETY OF CHEMISTRY

PAPER

E. Fadda and M. G. Nixon

The transient manifold structure of the p53 extreme C-terminal domain: insight into disorder, recognition, and binding promiscuity by molecular dynamics simulations

PCCP

PAPER

View Article Online
View Journal | View Issue

Check for updates

Cite this: *Phys. Chem. Chem. Phys.*, 2017, **19**, 21287

The transient manifold structure of the p53 extreme C-terminal domain: insight into disorder, recognition, and binding promiscuity by molecular dynamics simulations†

E. Fadda * and M. G. Nixon

The p53 tumour suppressor is a transcription activator that signals for cell cycle arrest and apoptosis. In its active form p53 is a tetramer, with each monomer organised in domains with different degrees of structural stability, ranging from the well folded DNA-binding domain (DBD) and tetramerization domain (TET), to the intrinsically disordered transactivation domain (TAD), and extreme C-terminal domain (CTD). Compared to all other domains, the structure/function relationship of the p53-CTD within the full-length p53 tetramer is still poorly understood due to its high degree of conformational disorder. Meanwhile, the structure of p53-CTD-like peptides has been well characterized when in complex with a variety of receptors, where, as other intrinsically disordered regions (IDR), it adopts specific, while diverse, conformations. Receptor-specific folding is likely to occur upon binding, either from a random coil, or as a result of an initial recognition of a pre-formed structural motif, known as molecular recognition feature (MoRF), selected by the receptor within the conformational ensemble of the IDP in solution. In this latter case, MoRFs act as nucleation sites, favouring the initiation of the folding process within the binding site. In this work we show the results of over 20 μs of cumulative molecular dynamics (MD) simulations of a 22 residue peptide unbound in solution with sequence corresponding to the p53-CTD 367–388 section. Such extensive sampling allowed us to identify and characterize the structure of specific sets of minimal structural MoRFs within the p53-CTD peptide conformational ensemble at equilibrium. These motifs are short, involving only 3 to 4 residues, and specifically localized within the peptide sequence. Corresponding patterns of secondary structure propensity along the p53-CTD sequence are also predicted by disorder prediction calculations. Based on these findings we discuss how the structural complementarity of specific minimal structural MoRFs to the binding site of different receptors could regulate the p53-CTD binding promiscuity.

Received 17th April 2017,
Accepted 2nd June 2017

DOI: 10.1039/c7cp02485a

rsc.li/pccp

Published on 02 June 2017. Downloaded by Maynooth University on 10/30/2019 9:44:44 AM.

Introduction

Conformational disorder is a common trait in a large part of proteins encoded in the human genome.[1–3] In contrast with the classic 'structure-function' relationship view, according to which protein function is inextricably related to a specific and stable native structure, the functional role of protein disorder is manifold,[4–6] crucial in regulatory and signalling pathways,[1,3,7] as well as conferring key mechanical properties to proteins.[8,9] The interaction between structured and intrinsically disordered regions (IDR) seems to follow also different principles relative to the formation of obligate protein–protein complexes.[5,7,10] Indeed, in obligate complexes the counterparts have highly complementary and preformed contact areas,[11,12] and only minor conformational changes occur upon binding.[13] IDR are not only structureless in solution, but often display broad specificity towards a set of different receptors.[1,7] Moreover, their bound conformation can vary widely in function of the specific receptor they are in complex with.[2,5,11,14,15] Contact areas between IDR and their receptors are generally small,[16] contributing to lower the binding affinity, thus favouring transient and reversible binding.[15–17] One prime example of this behaviour is given by the p53 extreme C terminal domain (CTD), a 30 residue highly basic IDR of the p53 tumour suppressor. The p53-CTD has been shown to have a negative regulatory control of the p53 DNA-binding activity,[18–20] with phosphorylation or deletion of the p53-CTD region resulting in a constitutively DNA-binding active p53 molecule.[21] The p53-CTD is also highly targeted for post-translational modifications, which modulate its DNA-binding activity.[21,22] The p53-CTD is poorly structured in solution,[23,24] but it adopts a variety of stable secondary structures when bound

Department of Chemistry, Maynooth University, and Hamilton Institute, Maynooth University, Maynooth, Kildare, Ireland. E-mail: elisa.fadda@nuim.ie

† Electronic supplementary information (ESI) available. See DOI: 10.1039/c7cp02485a

Published on 02 June 2017. Downloaded by Maynooth University on 10/30/2019 9:44:44 AM.

to different receptors, ranging from α helices, to coils.[14,23,25] Recognition and binding could follow these different mechanistic scenarios, (a) a non-specific ''encounter complex'' is initially formed between peptide and receptor, followed by an induced fit phase, where the progressive setting of specific interactions drives folding, or (b) the peptide unbound in solution can access its bound fold, which is selected and bound by the target receptor, *i.e.* by conformational selection, or (c) a mechanism in between these two scenarios, whereby conformational selection and induced fit both play a role.[26,27] In the specific case of the p53-CTD, it has been determined that the folded conformations are stable only when the peptide is bound.[23,28,29] Intricate balances regulating recognition and binding are not unusual for IDPs.[23,28–35] Because of the high degree of intrinsic disorder, structural investigations of the unbound p53-CTD in solution have not been particularly informative in terms of subtleties in its residual secondary structure.[19] Nevertheless, this does not necessarily mean that there isn't any residual secondary structure. Indeed, single molecule fluorescence experiments have shown that peptides previously characterized as completely disordered by ensemble approaches, do stochastically switch between subsets of conformations.[36] As a specific class of single molecule experiments, molecular simulation methods have been shown high potential in describing the structure, dynamics, and energetics of structural disorder at the atomistic level of detail[33,37–40] and results of such studies can provide important information that can be used to guide targeted experiments.

Recent work strongly suggests that recognition and binding mechanisms of IDPs depends on their intrinsic secondary structure propensity, dictated by the peptide/protein sequence.[34] In this work we used extensive molecular dynamics (MD) simulations to analyse at the atomistic level of detail the conformational propensity of the unbound p53-CTD in solution and to investigate the structure and potential roles of transiently stable structural motifs in recognition and binding.[41] Indeed, these motifs could function as molecular recognition features (MoRFs),[7,42] by being selectively targeted by different receptors, thus working as nucleation sites for the completion of the folding by induced fit.[1,3,4,41] Also, an intrinsic propensity to form minimal structural motifs that can be specifically recognized by different receptors would explain how the p53-CTD conformational disorder supports its binding promiscuity.[43]

The analysis of over 20 μs of cumulative MD simulation trajectories of a 22 residue peptide free in solution, with sequence corresponding to the p53-CTD $_{367}$SHLKSKKGQSTSRHKKLMFKTE$_{388}$ segment, suggest that the p53-CTD conformational disorder includes in addition to random coils, also structures containing specific, relatively stable, localized and reoccurring short secondary structure motifs, which encompass stretches of 3 to 4 residues in case of helical turns. The results also show that the MoRFs we identified occur with higher probability in the C-terminal half of the peptide, while the N-terminal half remains mostly disordered. These findings are in agreement with structure/disorder prediction tools, namely s2D[44] and PONDR-VL-XT[45] that show a different degree of structural propensity along the p53-CTD sequence. The identification of such distinct motifs within the disordered ensemble suggests that the p53-CTD may exert its broad binding specificity through minimal structural MoRFs, which are specifically selected and bound by different receptors, fitting within a broader framework of the conformational selection theory.[26,27] We discuss a potential MoRF-based recognition and binding mechanism in the case of the p53-CTD peptide in complex with the Ca^{2+} bound S100B(ββ) dimer,[23] and with sirtuin Sir2.[25]

Computational method

A 22 residue peptide corresponding to the 367–388 aa section of the *H. sapiens* p53-CTD was built in a fully extended conformation with the molecular builder tool in Maestro v.9.7.[46] N- and C-termini were capped with ACE and NME residues, respectively. The fully extended peptide, measuring 8.4 nm, was centred in a truncated dodecahedral simulation box sized so that the minimum distance between the peptide and the box sides would not be lower than 1.2 nm. The total charge of +6 was neutralized with the addition of Cl^- counterions. Because the aim of this work is to determine the conformational propensity of the peptide in function of its sequence, the effect of ionic strength in physiological conditions has not been addressed. Protein atoms and counterions were represented with AMBER-99SB-ILDN parameters,[47] while TIP4P-Ew[48] was chosen as water model. Because the choice of force field parameters can greatly affect the reliability of the simulations, the search for an optimal force-field selection for poorly structured and unstructured proteins is a topic very much under scrutiny nowadays.[37,38,49–53] The large number of existing protein/solvent parameters combinations and the complex behaviour of proteins in terms of their intrinsic structural propensity or lack-there of, makes the designation of an absolute best choice quite difficult.[37,38,51–54] Our selection of parameters was based on the proven reliability and robustness of the AMBER-99SB-ILDN/TIP4P-Ew force field in the simulation of structured proteins,[37,52] in view of its satisfactory performance in the prediction of residual secondary structure in disordered peptides.[38] It should be noted that AMBER-99SB-ILDN, when in combination with TIP3P, has been shown to underestimate chain dimensions[37,38] and has shown slower dynamics relative to other force fields combinations.[51] Convergence of our simulations has been verified by monitoring, (a) the average backbone RMSD values, calculated relative to 2 distinct and highly populated MoRFs we identified, namely an asymmetric β sheet hairpin (cluster 1, MD1) and a conformer containing a 3_{10} helical turn located at $_{376}$STS$_{378}$ (cluster 2, MD3), and (b) the corresponding backbone RMSD running averages, and (c) the RMSD average correlation values (RAC).[55] These data sets are included as Supplementary Material and shown in Fig. S1 (ESI†), panels (a)–(c), respectively.

Bond lengths with hydrogen atoms were constrained with the LINear Constraint Solver (LINCS) algorithm. Long range electrostatic interactions were treated with Particle Mesh Ewald (PME), with a 1.2 nm cutoff. van der Waals interactions were also cutoff at 1.2 nm. All MD simulations were run with version 4.6.3 of GROMACS.[56] Two additional simulations, each of 2 μs

 This journal is © the Owner Societies 2017

Published on 02 June 2017. Downloaded by Maynooth University on 10/30/2019 9:44:44 AM.

production, were run starting with the p53-CTD peptide in the helical conformation seen when in complex with S100B(ββ) (PDBid 1dt7). The following protocol was used to set-up and run all simulations in this work. An initial energy minimization of 500 k steps of steepest descent were used to prepare the system for the equilibration phase. During the minimization only the positions of the solvent molecules, counterions, and hydrogen atoms was left unconstrained. 500 ps of equilibration in the *NVT* ensemble and subsequently 500 ps in the *NPT* ensemble followed, with a target temperature of 300 K and pressure of 1 bar. Following this stage, a 1 ns equilibration was run with the backbone atoms restrained (C–N–Cα) and the sidechain atoms free. All restraints were then released for 5 ns and a 100 ns production run was recorded for analysis. From this trajectory, snapshots were collected every 10 ns. The 10 uncorrelated peptide structures were removed from their original simulation box and placed in a truncated dodecahedral simulation box of 8.4 nm sides, sized to leave enough space to accommodate largely extended conformations that may occur during the simulation. The minimization and equilibration protocol described above was repeated for all 10 systems. Production of every trajectory was extended to 2 μs, for a total simulation time of 20 μs. The choice of running a series of independent MD simulations instead of enhanced sampling schemes was based on the fact that we wanted to monitor the dynamic interconversion of MoRFs with the aim of identifying an interconversion coordinate, if possible. The clustering analysis was performed with the gromos algorithm[57] implemented in the g_cluster tool in GROMACS, with a cutoff of 0.45 nm. Such large cutoff value is necessary due to the highly dynamic nature of the N-terminal tail of the peptide, and it was found as the minimum cutoff value that allows us to separate and group significant secondary structure motifs. The secondary structure analysis was done on the 10 highest populated clusters obtained from each MD trajectory. These clusters are representative of a minimum of 63% of all conformations accessible, in case of high conformational dynamics, up to a complete coverage of 100%, in case of the formation of stable secondary structure motifs. MoRF populations for each MD trajectory were calculated based on clusters populations, with counts of 1 ns per frame. Populations over the cumulative 20 μs simulation were estimated as sums of the populations during the single trajectories, where the error bars corresponds to the standard deviation. Secondary structures were assigned according to the STRIDE online tool.[58] Image rendering, structural alignments, and distance analysis was done with PyMOL v.1.6.

Results

The conformational dynamics of a 22 residue peptide, corresponding to the Ser 367 to Glu 388 section of the p53-CTD, was analyzed with molecular dynamics (MD) for a cumulative time of 20 μs. The simulation is an aggregate of 10 separate trajectories, started from uncorrelated snapshots, selected from a 100 ns trajectory started from a fully extended backbone conformation. In addition to these 10 separate trajectories, we ran 2 separate, 2 μs long, trajectories, both started from the helical conformation of p53-CDT peptide when in complex with S100B(ββ)[23] (PDBid 1dt7). These calculations were run to assess the stability of the helix in solution and the residual helicity retained in function of the starting structure.

The p53-CTD peptide size was analysed in terms of its radius of gyration (R_g). The average R_g calculated over the combined 20 μs trajectories is 0.97 ± 0.07 nm. This value is larger than the R_g value 0.73 nm, predicted for a 22 residue random-flight polymer with link distances of 0.38 nm, representative of a random coil behaviour, but smaller than 1.01 ± 0.04 nm, the R_g measured for the Ace-$(AAKAA)_4$-GY-NH_2 peptide, containing well structured α helical motifs.[59] It should be noted that the Amber-99SB-ILDN has been shown to produce an overall more compact ensemble relative to experiment.[37,38] The R_g plots obtained from the 10, 2 μs long, trajectories are shown in Fig. S3 (ESI†). The trends describe a highly dynamic structure with recurrent signatures of relatively stable conformers. These correspond to localized and short secondary structure motifs, which we have classified through clustering analysis and are described in the sections below. The relative populations of the secondary structure motifs identified over the 20 μs of cumulative simulation time are shown in Fig. 1 and in Table 1. The highest populated structures contain β-bridges (27.3%), and β-sheet hairpins (25.7%). We also identified helical turns (13%), both α-helical, and 3_{10}. Random coils and turns, which contribute the most to the disordered character of the peptide, have a relative population of 15.2%. Interestingly, similar secondary structure motifs form recurrently throughout the dynamics, and are found to involve preferentially the same group of residues within the C-terminal half of the peptide.

β-Sheet hairpin motifs

As shown in the R_g plot in Fig. S3 (ESI†), the first MD trajectory (MD1) visits a relatively stable conformation between 400 ns and 1.5 μs. In this interval the peptide is in an asymmetric β-sheet conformation, shown in Fig. 2. The core of the β-sheet is held together by hydrogen bonds connecting Ser 376 and Phe 385, while the free N-terminal tail, spanning residues Ser 367 to Gly 374 ($_{367}$SHLKSKKG$_{374}$), is not tied into the hairpin and its dynamics determines the oscillations of the R_g value. The hairpin turn comprises His 380 and Lys 381. This conformation is the highest populated, stable for 55% of the simulation time, or more than half of the MD1 trajectory. The structural alignment of all the β-sheet motifs found during all the other simulations shows that this particular asymmetric hairpin conformation is the highest populated type of β-sheet, also present in MD6 with 63.4% population over 2 μs, MD7 with 17.7% population, MD8 with 89.0% population, and MD10 with population of 23.0%. Two other slightly different, and lower populated, β-sheet conformations have been identified, one in MD5, with a population of 3.8% over 2 μs, and the other in MD10, with a population of 4.4%. As shown in Fig. S3 in ESI,† these two conformers show the same asymmetry, but one (MD10) has a wider hairpin section, formed by Lys 381, Lys 382, and Leu 383, and the other one

Published on 02 June 2017. Downloaded by Maynooth University on 10/30/2019 9:44:44 AM.

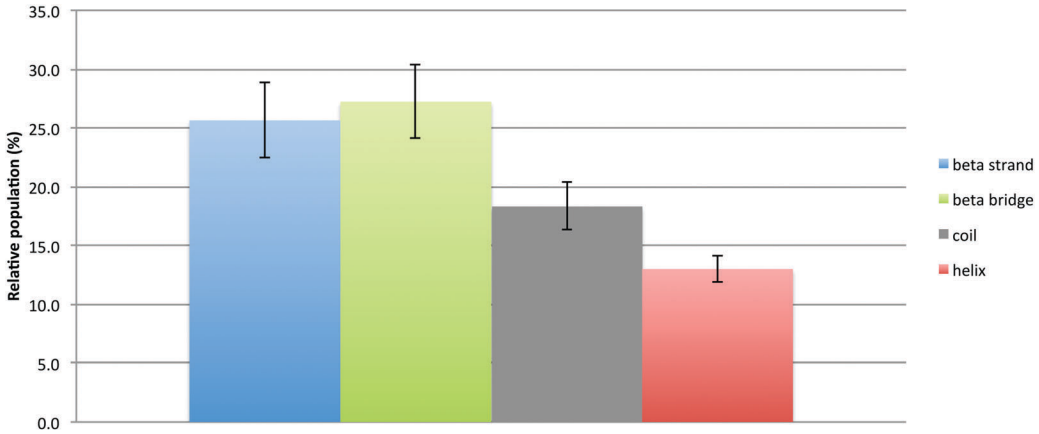


Fig. 1 Relative populations of secondary structure motifs identified over a cumulative 20 μs MD simulation of the 22 residue p53-CTD peptide.

Table 1 Relative populations (%) of the secondary structure motifs identified during each trajectory. MD 1–10 are trajectories started from a common fully extended peptide, see Computational method section, while Helix 1–2 MD indicate trajectories started from the α helical conformation from the complex with S100B(ββ) (PDBid 1dt7). Populations are calculated over 2 μs and account for the 10 highest populated clusters. The total reflects the populations over 20 μs cumulative sampling, where the standard deviation is indicated in brackets

Trajectory	β-Strands	$\alpha/3_{10}$ helices	β-Bridges	Coil/turns
MD 1	55.3	7.7	8.9	13.4
MD 2	0.0	19.1	0.0	53.4
MD 3	0.0	27.8	24.4	17.3
MD 4	0.0	0.0	81.4	14.5
MD 5	3.8	19.9	9.4	14.4
MD 6	63.4	7.1	26.5	4.7
MD 7	17.7	21.1	83.1	0.0
MD 8	89.0	0.0	8.9	2.2
MD 9	0.0	2.9	0.0	57.6
MD 10	27.5	24.8	30.1	6.0
Total (over 20 μs)	**25.7(3.2)**	**13.0(1.1)**	**27.3(3.1)**	**18.3(2.0)**
Helix 1 MD	0.0	52.7	0.0	30.7
Helix 2 MD	0.0	11.7	41.4	29.0

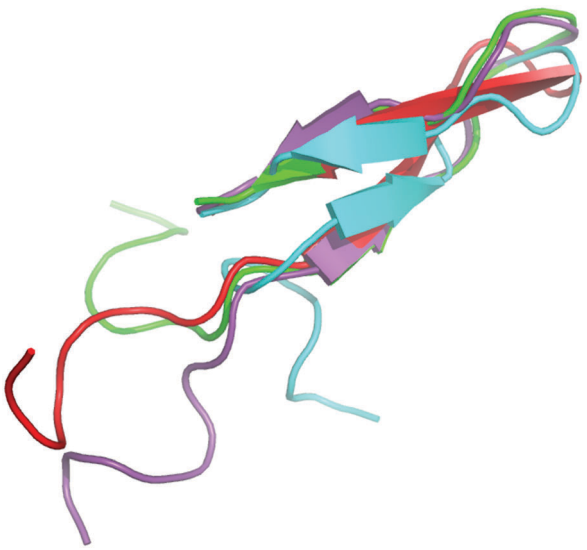


Fig. 2 Examples of the β-sheet motif identified through the clustering analysis of the 20 μs MD simulation of the p53-CTD. The conformations visited during MD 1, MD 7, and MD 8 (cluster 1 and 3), are shown in cyan, red, green, and purple, respectively. The flexible tail corresponds to the stretch between Ser 367 and Gly 374.

(MD5) with a slightly different hydrogen bonding pattern relative to the highest populated motif, connecting not only Ser 376 to Phe 385, but also Ser 378 to Met 386. An RMSD matrix obtained through sequence alignment of the 15 residues stretch between Gly 375 and Glu 388, followed by structural alignment of all the β-sheet motifs, is shown in Table S2 in the ESI.†

β-Bridges containing motifs

As shown in Fig. 1, structures containing β-bridges are the highest populated over the 20 μs MD. Because of the degree of conformational flexibility a single hydrogen bond allows, structures containing β-bridges can be quite different, ranging from elongated narrow hairpins, to globular folds containing one or two β-turns, see Fig. S5 (ESI†). The most stable hairpins that contain β-bridges are structurally similar to asymmetric hairpins described in the previous section. In fact, these β-bridge hairpins also present along dynamic tail that comprises residues Ser 367 to Gly 374 ($_{367}$SHLKSKKG$_{374}$) and also have His 380 and Lys 381 at the hairpin turn. The largest group of stable, narrow, asymmetric hairpins was visited during MD7 with 76.7% population over 2 μs MD, see Fig. S4 (ESI†). The structure of the highest populated hairpins containing β-bridges and β-sheet hairpins are structurally very similar; they fall in different categories as the MoRFs structure classification we used is based on the STRIDE definition of the clusters middle structure. Indeed, clusters of narrow hairpins with a β-bridge often also contain β-sheets and *vice versa*.

3_{10} and α helical turns

Over the 20 μs MD we were able to distinguish 3 significantly populated short helical motifs, either 3_{10} or α single helical turns. Representative structures are shown in Fig. 3. The highest populated helical MoRF involves a 3 residue segment, between

 This journal is © the Owner Societies 2017

Published on 02 June 2017. Downloaded by Maynooth University on 10/30/2019 9:44:44 AM.

Ser 376 and Ser 378 ($_{376}$STS$_{378}$), with a relative population of 6.1% (±0.3) over the cumulative 20 μs. Two helical motifs are equally populated; one is located at the N-terminus end of the peptide, stretching across Lys 370 to Lys 372 ($_{370}$KSK$_{372}$) with a relative population of 2.4% (±0.2) over 20 μs, and the other is located at the C-terminus end, between Lys 381 and Leu 383 ($_{381}$KKL$_{383}$), with a relative population of 2.3% (±0.8). Although we found that the single helical turn motifs are in general less stable overtime than the asymmetric β-sheet or β-bridge-containing hairpins, we observed that the highest populated helical turns are stable between 250 and 450 ns, see Fig. S6 (ESI†).

Stability of the S100B(ββ)-bound conformation in solution

We ran 2 additional 2 μs trajectories to determine the stability in solution of the α/3_{10} helical structure of the 22 residue p53-CTD peptide when in complex with the S100B(ββ) dimer.[23] According to the STRIDE classification, this structure is α helical from Thr 377 to Met 384, and 3_{10} helical from Phe 385 to Thr 387. As shown by the NMR ensemble[23] (PDBid 1dt7), large part of the N-terminal half of the peptide, *i.e.* from Ser 367 to Ser 376, is unbound and highly dynamic. Although the helical structure of the peptide unfolds quite readily within the first 2 ns of both trajectories, the secondary structure analysis in Table 1 shows that a higher degree of helicity remains in one of the two 2 μs trajectories, namely in Helix 1 MD, relative to all other trajectories originated from the common fully extended starting structure. A high degree of helicity in the unbound p53-CTD was described in a recently published computational work.[60] Such strong conformational propensity is inconsistent with circular dichroism data[24,60] and could be due not only to force field limitations, but also to the choice of starting structures derived from the S100B(ββ) bound conformation.[60] In 30.7% of the helical structures, the residual helicity spans the $_{381}$KKL$_{383}$ stretch. In all other cases, single helical turns are observed in the disordered, and unbound N-terminal tail, often in addition to the helical turn at $_{381}$KKL$_{383}$. As shown in Table 1, also during the Helix 2 MD simulation, a residual helical character remains, however much less predominant than in the case of Helix 1 MD. Indeed, the highest populated cluster corresponds to an asymmetric hairpin structure with a free N-terminal tail (Ser 367 to Lys 373), similar to the β-bridge and β-sheet hairpin motifs observed during the 20 μs simulation MD 1 to 10, described in the previous sections. Average R_g values are 0.996 nm for Helix 2 MD and 1.059 nm for Helix 1 MD, see Fig. S7 (ESI†).

Discussion

The extreme C-terminus of the p53 tumor suppressor (p53-CTD) is a 30 residue long conformationally disordered,[23] highly alkaline region, responsible for regulating the p53 DNA binding activity.[18–21] This intrinsically disordered region (IDR) is also highly targeted for post-translational modifications, which modulate its DNA binding activity.[18,19,22,23,61] The p53-CTD binds numerous receptors, adopting significantly different conformations when bound.[14,23,25] In this work we used extensive sampling *via* MD simulations to analyze the conformational propensity of a 22 residue peptide, bearing all the binding determinants of the p53-CTD,[23,62–64] while unbound in solution. Our objective was to characterize at the atomistic level of detail the dynamic nature of the p53-CTD peptide conformational ensemble at equilibrium and to search for structural distinctive elements, or Molecular Recognition Features (MoRFs), that could be specifically selected and bound by different receptors, initiating receptor-specific folding patters. As shown in Fig. 1, we have identified a set of distinct structural MoRFs, significantly populated over 20 μs of cumulative sampling. These include β-sheet and β-bridges-containing asymmetric hairpins, conformations with 3_{10} and α single helical turns, as well as coils and turns-containing structures, which account for the disordered nature of the peptide. An analysis of the size of the peptide in terms of R_g values in relation to the relative populations of the different structural MoRFs is shown in Fig. 4. The most compact, and highest populated, conformations correspond to asymmetric hairpin structures, examples of which are shown in Fig. 2, while the most extended ones correspond to disordered coils.

A comparison between the secondary structure propensity per residue calculated over the 20 μs of MD simulation and the disorder predictions obtained with the s2D[44] and with the PONDR-VL-XT[45] tools is shown in Fig. 5. Provided that the disorder scores obtained from the structure prediction tools are

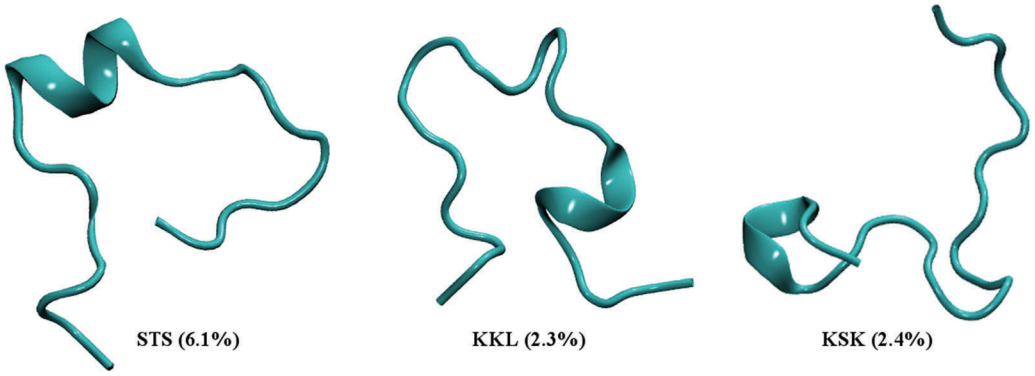


Fig. 3 Short helical motifs identified during the 20 μs MD simulations. The labels indicate the group of residues where the helical turn is centred, and the minimum stretch for a 3_{10} turn, while the relative populations over 20 μs are indicated in brackets.

Published on 02 June 2017. Downloaded by Maynooth University on 10/30/2019 9:44:44 AM.

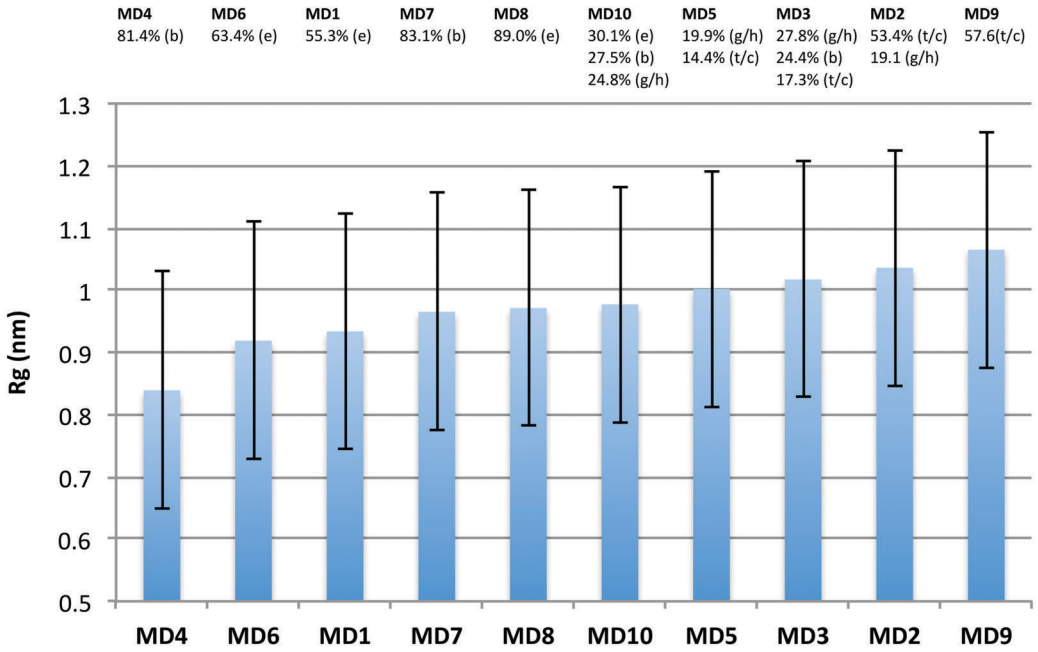


Fig. 4 Average R_g values (nm), calculated over the 10, 2 μs trajectories, ordered from smaller to larger values. Above each bar are indicated the largest percentages of secondary structures identified during each trajectory, with (e) indicating β-sheet motifs, (b) β-bridges, (g h^{-1}) $3_{10}/\alpha$ helical single turns, and (t/c) turns and coils.

not numerically comparable to the secondary structure propensity values calculated from the MD simulations, or to each other, our analysis shows a good agreement with the s2D data, which predict a decrease in disorder (<70% disorder score) in the $_{377}$TSRHKKLMFKT$_{387}$ segment, where we also observe the highest propensity for secondary structure. The PONDR-VL-XT prediction is also in agreement with our data, showing a smoother decrease in disorder from N- to C-terminus, with a significant decrease for a slightly shorter sequence range relative to the s2D data, namely $_{380}$HKKLMFKTE$_{388}$. As shown in Fig. 5, the MD data provide a rationalization for this decrease in disorder in the C-terminal half of the peptide, by showing a higher propensity for the formation of β-sheet asymmetric hairpins with a disordered N-terminal tail, and single helical turns. The propensity to form short helical motifs is also detected in the mostly disordered N-terminal half of the peptide, a trend mirrored by a slight decrease in disorder predicted by the s2D tool, see Fig. 5, panel (a).

The identification of these minimal structural motifs, or MoRFs, provides a rationale that can explain the binding promiscuity of p53-CTD, or its specificity towards multiple receptors. Our working hypothesis is that the p53-CTD receptors have different binding affinities for the structural MoRFs accessible at equilibrium, because of their 'preformed' 3D spatial arrangements, which provides structural complementarity to different binding site. As the MoRFs constitute minimal structural motifs, the MoRF-receptor structural complementarity provides only on a few protein/peptide specific contacts, resulting in a relatively low binding enthalpy contribution. For instance, when in complex with the S100B(ββ) dimer, the p53-CTD peptide is in a helical conformation.[23] Within our the minimal MoRF-recognition framework, the S100B(ββ) dimer selects one (or more) of the single helical turn-containing MoRFs found in the p53-CTD conformational ensemble to form an initial recognition (or nucleation) complex. Completion of the folding, with a corresponding increment of the binding affinity, will proceed by induced fit.[26] The completion of the helical motif after binding is in agreement with the observation that in the absence of stabilizing tertiary interactions, α-helices rarely persist in isolation,[59] and it is also supported by the instability of the S100B(ββ)-bound helical conformation of the p53-CTD peptide when unbound in solution, see Table 1.

As shown in Fig. 6, an example of a potential recognition complex can be built by structural alignment of the backbone atoms of the $_{376}$STS$_{378}$ helical turn MoRF, the highest populated helical motif over the cumulative 20 μs, onto the bound peptide conformation. The structure of the S100B(ββ)/p53-CTD NMR complex[23] reveals sets of specific ligand-receptor interactions, which include a hydrogen bond between Ser 376 of p53-CTD and Glu 45 of S100B(ββ), an salt-bridge between Arg 379 of p53-CTD and Glu 49 of S100B(ββ), and the insertion of Leu 383 into the hydrophobic binding groove. As shown in Fig. 6, these interactions are conserved in the putative recognition complex and retained during a 100 ns MD structure relaxation run. Notably, the disordered character of the unbound N-terminal tail of the peptide highlighted in the NMR structure[23] is also well reproduced.

Another potential example of MoRF-based conformational selection can explain the binding of a p53-CTD peptide with sirtuin Sir2 (PDBid 2h2f).[25] The only resolved residues of the p53-CTD peptide in the complex are the one directly in contact with the Sir2 binding site, namely $_{378}$SRKKLM$_{383}$. While most of

 This journal is © the Owner Societies 2017

Published on 02 June 2017. Downloaded by Maynooth University on 10/30/2019 9:44:44 AM.

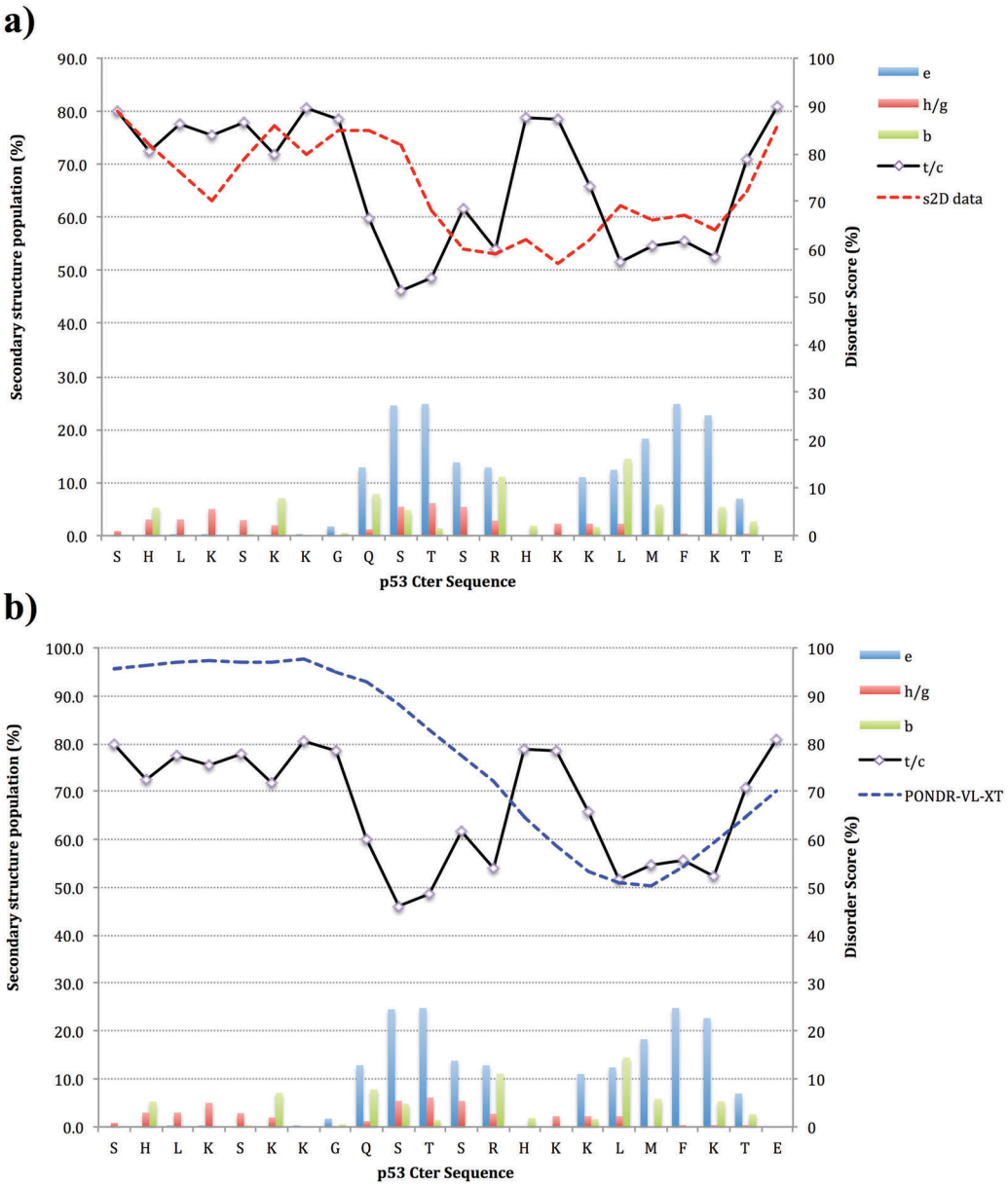


Fig. 5 Comparison of the conformational propensity per residue calculated from the 20 μs MD simulations to the disorder prediction tools s2D on panel (a), and PONDR-VL-XT on panel (b). The legends on the right-hand side of the graphs indicate specific secondary structure motifs, namely β sheets (e), α (h) and 3_{10} helices (g), β bridges (b), turns (t) and coils (c). Secondary structure assignments have been done with STRIDE.[58]

the contact between Sir2 and its target peptide involve mostly backbone atoms,[25] as shown in Fig. 7, when in complex with the unmodified p53-CTD peptide, a few significant specific contacts can be highlighted. More specifically, Lys 381 of p53-CTD is in a salt bridge with the Gly 163 backbone carbonyl of Sir2, while the sidechain of Lys 382, targeted by acetylation, protrudes into the Sir2 binding site, with Phe 162 and His 116 flanking the aliphatic side chain, and the amino group bound to a water molecule.[25] Based on the structural data available, the transition from conformational disorder to order upon binding appears to be less significant in this case relative to the S100B(ββ). Nevertheless, if we consider the interactions between the Sir2 and Lys 381 and 382 of p53-CTD as potential recognition contacts, we have 2 examples, shown in Fig. 7, of potentially recognized MoRFs in coil/turn conformation, with an optimal orientation of the Lys sidechains for initial recognition.

To our knowledge there are no structures of complexes with the unmodified p53-CTD in an asymmetric β-sheet conformation; nevertheless, we are currently investigating the recognition of the highest populated p53-CTD MoRF, namely the β-sheet or β-bridge-containing asymmetric hairpin, by β-sheet structured binding sites, such as in the PCL1-PHD1 domain.[65]

Because of their relatively low populations and frequent interconversion, the p53-CTD peptide structural MoRFs will be extremely difficult, if not impossible, to characterize experimentally.

Published on 02 June 2017. Downloaded by Maynooth University on 10/30/2019 9:44:44 AM.

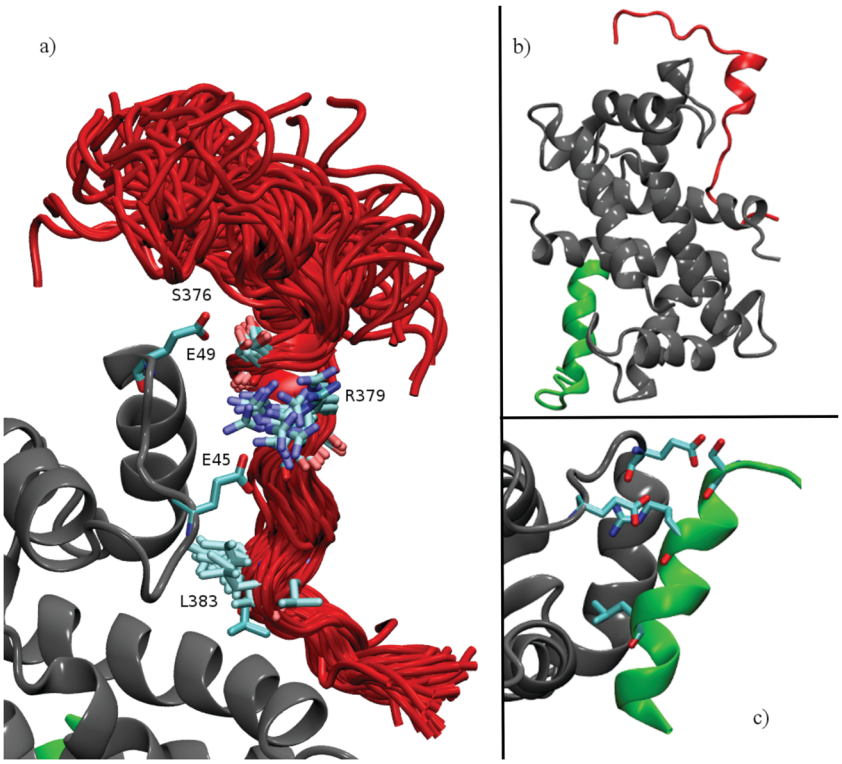


Fig. 6 (Panel a) Close up on specific contacts the helical MoRF p53-CTD peptide (red) makes during the 100 ns MD when in one of the potential recognition complexes with S100B(ββ) (grey). Shown from top to bottom, the hydrogen bond interaction between Ser 376 of p53-CTD and Glu 49 of S100B(ββ), the salt bridge between Arg 379 of p53-CTD and Glu 45 of S100B(ββ), and the interaction of Leu 383 inserted in the hydrophobic binding groove of S100B(ββ). (Panel b) Structure of the recognition complex built by structural alignment of a peptide containing the STS helical (red) turn on to the p53-CTD bound conformation. (Panel c) Set of contacts highlighted in panel (a) shown for the bound conformation of the p53-CTD peptide (green) in complex with S100B(ββ), PDBid 1dt7.

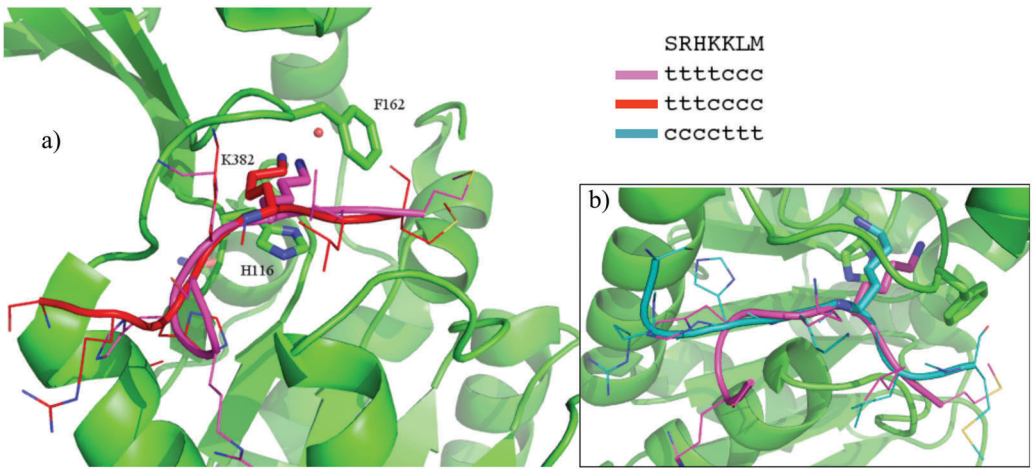


Fig. 7 Potential recognition complexes between the p53-CTD peptide (in purple) and Sir2Tm (in green), PDBid 2h2f, obtained by structural alignment of MoRFs identified through clustering analysis of 20 μs MD trajectories. In panel (a) alignment of the middle structure from cluster 8 of MD5 (in red) with a RMSD based on 4 Ca atoms of 0.4 Å, in panel (b) alignment of the middle structure from cluster 8 of MD9 (in cyan) with a RMSD based on 4 Ca atoms of 0.3 Å. Peptide sequence and respective secondary structures assignments (STRIDE) are indicated in the legend.

Indeed, existing NMR[23] and CD[60] data show that the p53-CTD section is highly disordered when unbound in solution. Nevertheless, based on the information discussed in this work, experimental support for the MoRF-driven molecular recognition mechanism could be obtained by biasing the conformational ensemble towards specific structural MoRFs by means, for example, of stapling the peptide.[66] More specifically, based on the potential recognition discussed earlier in this section of the $_{376}$STS$_{378}$ helical turn by S100B(ββ) as the initial, low affinity recognition complex, a suitably placed aliphatic chain staple

 This journal is © the Owner Societies 2017

Published on 02 June 2017. Downloaded by Maynooth University on 10/30/2019 9:44:44 AM.

would enhance the conformational propensity of the helical MoRF, thus its population, without affecting sequence integrity. Changes in relative conformers populations would affect binding kinetics. Furthermore, the role of MoRFs in p53-CTD recognition and binding could also be tested by introducing mutations that suppress specific structural motifs. Indeed, in the case of the small disordered protein PUMA, an Ala-Gly scanning scheme was put in place to suppress helicity without affecting residual structural propensity in solution.[30] Binding kinetics showed that the mutations do not affect $k+$, suggesting that folding occurs by an induced fit-driven mechanism.[30] As a proof of principle, based on the recognition mechanism proposed earlier for the $_{376}STS_{378}$ 3_{10} helical turn MoRF by S100B(ββ), mutations of Thr 377 and/or Ser 378 would affect conformational propensity without compromising specific contacts with the receptor. Work in this direction is currently underway.

Conclusions

In this work we have used extended conformational sampling through conventional MD simulations to determine the degree of residual secondary structure within the conformational disorder at equilibrium of a 22 residue peptide, corresponding to the 367–388 region of the p53 C-terminal domain (p53-CTD). This peptide contains all binding determinants of the p53-CTD within the active p53 tumour suppressor.[23,25] Clustering analysis of the MD trajectories, accounting for a cumulative time of over 20 μs, show the p53-CTD peptide has a high conformational flexibility, but also a distinct propensity for the formation of specific and short structural motifs that encompass 3 to 4 residue at most. Furthermore, a *per* residue analysis of the conformational propensity along the p53-CTD peptide shows that these structural motifs are localized along the sequence, involving specific groups of residues. Localization of the structural MoRFs makes the p53-CTD C-terminal half less disordered than the N-terminal half. This observation is also in agreement with disorder predictions obtained with the s2D[44] and PONDR-VL-XT[45] secondary structure and disorder prediction tools. We propose that the functional role of these minimal structural molecular recognition features (MoRFs) is to confer to the p53-CTD binding specificity towards different receptors, whereby each receptor would have a higher affinity for a specific MoRF due to 3D structural complementarity, based on a reduced number of contacts, relative to the final, bound conformation. Molecular recognition through selection of specific MoRFs, would lead to the completion of folding through induced fit. This mechanism is consistent with the molecular recognition proposed for other IDP systems,[34] although it does not necessarily preclude access to other recognition and binding pathways,[30,34] that could in principle coexist.

Authors contributions

EF designed research; EF and MN performed research; EF and MN analyzed the data; EF wrote the paper.

Acknowledgements

The authors gratefully acknowledge Dr Sarah Rauscher for insightful discussions and the Irish Centre for High-End Computing (ICHEC) for the generous allocation of computational resources. MN acknowledges the John and Pat Hume Scholarship Programme, Maynooth University, for financial support.

References

1 M. M. Babu, R. van der Lee, N. S. de Groot and J. Gsponer, *Curr. Opin. Struc. Biol.*, 2011, **21**, 432–440.
2 R. van der Lee, M. Buljan, B. Lang, R. J. Weatheritt, G. W. Daughdrill, A. K. Dunker, M. Fuxreiter, J. Gough, J. Gsponer, D. T. Jones, P. M. Kim, R. W. Kriwacki, C. J. Oldfield, R. V. Pappu, P. Tompa, V. N. Uversky, P. E. Wright and M. M. Babu, *Chem. Rev.*, 2014, **114**, 6589–6631.
3 P. E. Wright and H. J. Dyson, *Nat. Rev. Mol. Cell Biol.*, 2015, **16**, 18–29.
4 P. Tompa, *FEBS Lett.*, 2005, **579**, 3346–3354.
5 E. Hazy and P. Tompa, *ChemPhysChem*, 2009, **10**, 1415–1419.
6 P. Tompa, *Curr. Opin. Struct. Biol.*, 2011, **21**, 419–425.
7 H. J. Dyson and P. E. Wright, *Nat. Rev. Mol. Cell Bio.*, 2005, **6**, 197–208.
8 S. Rauscher and R. Pomès, *Adv. Exp. Med. Biol.*, 2012, **725**, 159–183.
9 S. Rauscher, S. Baud, M. Miao, F. W. Keeley and R. Pomès, *Structure*, 2006, **14**, 1667–1676.
10 R. Pancsa and M. Fuxreiter, *Iubmb Life*, 2012, **64**, 513–520.
11 J. Janin, R. P. Bahadur and P. Chakrabarti, *Q. Rev. Biophys.*, 2008, **41**, 133–180.
12 I. M. A. Nooren and J. M. Thornton, *EMBO J.*, 2003, **22**, 3486–3492.
13 L. Lo Conte, C. Chothia and J. Janin, *J. Mol. Biol.*, 1999, **285**, 2177–2198.
14 A.-S. Huart and T. R. Hupp, *BioDiscovery*, 2013, **8**, 5.
15 B. Meszaros, Z. Dosztanyi and I. Simon, *PLoS One*, 2012, 7.
16 J. R. Perkins, I. Diboun, B. H. Dessailly, J. G. Lees and C. Orengo, *Structure*, 2010, **18**, 1233–1243.
17 I. M. A. Nooren and J. M. Thornton, *J. Mol. Biol.*, 2003, **325**, 991–1018.
18 J. Ahn and C. Prives, *Nat. Struct. Biol.*, 2001, **8**, 730–732.
19 O. Laptenko, I. Shiff, W. Freed-Pastor, A. Zupnick, M. Mattia, E. Freulich, I. Shamir, N. Kadouri, T. Kahan, J. Manfredi, I. Simon and C. Prives, *Mol. Cell*, 2015, **57**, 1034–1046.
20 A. C. Joerger and A. R. Fersht, *Annu. Rev. Biochem.*, 2008, **77**, 557–582.
21 T. R. Hupp, D. W. Meek, C. A. Midgley and D. P. Lane, *Cell*, 1992, **71**, 875–886.
22 L. J. Feng, T. X. Lin, H. Uranishi, W. Gu and Y. Xu, *Mol. Cell. Biol.*, 2005, **25**, 5389–5395.
23 R. R. Rustandi, D. M. Baldisseri and D. J. Weber, *Nat. Struct. Biol.*, 2000, **7**, 570–574.
24 O. D. Shahar, R. Gabizon, O. Feine, R. Alhadeff, A. Ganoth, L. Argaman, E. Shimshoni, A. Friedler and M. Goldberg, *PLoS One*, 2013, **8**, e78472.

Published on 02 June 2017. Downloaded by Maynooth University on 10/30/2019 9:44:44 AM.

25 M. S. Cosgrove, K. Bever, J. L. Avalos, S. Muhammad, X. B. Zhang and C. Wolberger, *Biochemistry*, 2006, **45**, 7511–7521.
26 P. Csermely, R. Palotai and R. Nussinov, *Trends Biochem. Sci.*, 2010, **35**, 539–546.
27 D. D. Boehr, R. Nussinov and P. E. Wright, *Nat. Chem. Biol.*, 2009, **5**, 789–796.
28 W. J. Allen, D. G. S. Capelluto, C. V. Finkielsein and D. R. Bevan, *J. Phys. Chem. B*, 2010, **114**, 13201–13213.
29 S. Kannan, D. P. Lane and C. S. Verma, *Sci. Rep.*, 2016, **6**, 23750.
30 J. M. Rogers, V. Oleinikovas, S. L. Shammas, C. T. Wong, D. De Sancho, C. M. Baker and J. Clarke, *Proc. Natl. Acad. Sci. U. S. A.*, 2014, **111**, 15420–15425.
31 K. Gunasekaran and R. Nussinov, *Biophys. J.*, 2005, **88**, 3a.
32 T. Hayashi, H. Oshima, S. Yasuda and M. Kinoshita, *J. Phys. Chem. B*, 2015, **119**, 14120–14129.
33 A. Toto, C. Camilloni, R. Giri, M. Brunori, M. Vendruscolo and S. Gianni, *Sci. Rep.*, 2016, 6.
34 M. Arai, K. Sugase, H. J. Dyson and P. E. Wright, *Proc. Natl. Acad. Sci. U. S. A.*, 2015, **112**, 9614–9619.
35 L. A. Metskas and E. Rhoades, *J. Mol. Biol.*, 2016, **428**, 2965–2977.
36 U. B. Choi, J. J. McCann, K. R. Weninger and M. E. Bowen, *Structure*, 2011, **19**, 566–576.
37 S. Piana, J. L. Klepeis and D. E. Shaw, *Curr. Opin. Struct. Biol.*, 2014, **24**, 98–105.
38 S. Rauscher, V. Gapsys, M. J. Gajda, M. Zweckstetter, B. L. de Groot and H. Grubmuller, *J. Chem. Theory Comput.*, 2015, **11**, 5513–5524.
39 Y. Chebaro, A. J. Ballard, D. Chakraborty and D. J. Wales, *Sci. Rep.*, 2015, **5**, 10386.
40 A. H. Mao, S. L. Crick, A. Vitalis, C. L. Chicoine and R. V. Pappu, *Proc. Natl. Acad. Sci. U. S. A.*, 2010, **107**, 8183–8188.
41 M. Fuxreiter, I. Simon, P. Friedrich and P. Tompa, *J. Mol. Biol.*, 2004, **338**, 1015–1026.
42 T. Mittag, L. E. Kay and J. D. Forman-Kay, *J. Mol. Recognit.*, 2010, **23**, 105–116.
43 P. Tompa, *THEOCHEM*, 2003, **666**, 361–371.
44 P. Sormanni, C. Camilloni, P. Fariselli and M. Vendruscolo, *J. Mol. Biol.*, 2015, **427**, 982–996.
45 P. Romero, Z. Obradovic, X. H. Li, E. C. Garner, C. J. Brown and A. K. Dunker, *Proteins*, 2001, **42**, 38–48.
46 Maestro, Schrödinger, LLC, New York, NY, 2014.
47 K. Lindorff-Larsen, S. Piana, K. Palmo, P. Maragakis, J. L. Klepeis, J. O. Dror and D. E. Shaw, *Proteins*, 2010, **78**, 1950–1958.
48 H. W. Horn, W. C. Swope, J. W. Pitera, J. D. Madura, T. J. Dick, G. L. Hura and T. Head-Gordon, *J. Chem. Phys.*, 2004, **120**, 9665–9678.
49 J. Huang, S. Rauscher, G. Nawrocki, T. Ran, M. Feig, B. L. de Groot, H. Grubmuller and A. D. MacKerell, Jr., *Nat. Methods*, 2017, **14**, 71–73.
50 J. Henriques, C. Cragnell and M. Skepo, *J. Chem. Theory Comput.*, 2015, **11**, 3420–3431.
51 F. Vitalini, A. S. J. S. Mey, F. Noe and B. G. Keller, *J. Chem. Phys.*, 2015, 142.
52 K. A. Beauchamp, Y. S. Lin, R. Das and V. S. Pande, *J. Chem. Theory Comput.*, 2012, **8**, 1409–1414.
53 S. Piana, A. G. Donchev, P. Robustelli and D. E. Shaw, *J. Phys. Chem. B*, 2015, **119**, 5113–5123.
54 K. Lindorff-Larsen, P. Maragakis, S. Piana, M. P. Eastwood, R. O. Dror and D. E. Shaw, *PLoS One*, 2012, **7**, e32131.
55 R. Galindo-Murillo, D. R. Roe and T. E. Cheatham, *BBA, Biochim. Biophys. Acta, Gen. Subj.*, 2015, **1850**, 1041–1058.
56 D. van der Spoel, E. Lindhal, B. Hess, G. Groenhof, A. E. Mark and H. J. Berendsen, *J. Comput. Chem.*, 2005, **26**, 1701–1718.
57 X. Daura, K. Gademann, B. Jaun, D. Seebach, W. F. van Gunsteren and A. E. Mark, *Angew. Chem., Int. Ed.*, 1999, **38**, 236–240.
58 M. Heinig and D. Frishman, *Nucleic Acids Res.*, 2004, **32**, W500–502.
59 B. Zagrovic, G. Jayachandran, I. S. Millett, S. Doniach and V. S. Pande, *J. Mol. Biol.*, 2005, **353**, 232–241.
60 S. Iida, T. Mashimo, T. Kurosawa, H. Hojo, H. Muta, Y. Goto, Y. Fukunishi, H. Nakamura and J. Higo, *J. Comput. Chem.*, 2016, **37**, 2687–2700.
61 Y. Xu, *Cell Death Differ.*, 2003, **10**, 400–403.
62 J. L. Avalos, I. Celic, S. Muhammad, M. S. Cosgrove, J. D. Boeke and C. Wolberger, *Mol. Cell*, 2002, **10**, 523–535.
63 E. D. Lowe, I. Tews, K. Y. Cheng, N. R. Brown, S. Gul, M. E. Noble, S. J. Gamblin and L. N. Johnson, *Biochemistry*, 2002, **41**, 15625–15634.
64 S. Mujtaba, Y. He, L. Zeng, S. Yan, O. Plotnikova, Sachchidanand, R. Sanchez, N. J. Zeleznik-Le, Z. Ronai and M. M. Zhou, *Mol. Cell*, 2004, **13**, 251–263.
65 G. L. Brien, E. Healy, E. Jerman, E. Conway, E. Fadda, D. O'Donovan, A. V. Krivtsov, A. M. Rice, C. J. Kearney, A. Flaus, S. S. McDade, S. J. Martin, A. McLysaght, D. J. O'Connell, S. A. Armstrong and A. P. Bracken, *Genes Dev.*, 2015, **29**, 2231–2243.
66 L. D. Walensky and G. H. Bird, *J. Med. Chem.*, 2014, **57**, 6275–6288.

 This journal is © the Owner Societies 2017