Intrinsically Disordered p53 and its Complexes Populate Compact Conformations in the Gas Phase

*Kevin Pagel,[1‡] Eviatar Natan,[2] Zoe Hall,[1] Alan R. Fersht,[2] Carol V. Robinson[1]**

[1] *Department of Chemistry, Physical & Theoretical Chemistry Laboratory, University of Oxford, South Parks Road, OX1 3QZ, Oxford, UK;* [2] MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 0QH, UK

[‡] *Current address: Fritz Haber Institute of the Max Planck Society, Department of Molecular Physics, Faradayweg 4-6, 10409 Berlin, Germany*

-Supporting Information-

Protein expression and purification

p53 constructs containing the core domain (C, CT and FL) were modified at four residues (M133L/V203A/N239Y/N268D)[1] and cloned into pET-24a with an N-terminal 6×His/lipoamyl domain tag and tobacco-etch-virus (TEV) protease cleavage-site. Each cloned plasmid was transformed into *Escherichia coli* BL21(DE3) and expressed and purified following published protocols.[2] The tetramerization domain (residues 325-356) was produced as described recently.[3] The cDNA clones were purchased from Geneservice Ltd. and transformed in *E. coli* Bl21 cells using a pRSET vector without any tag. Cells were grown and induced at 37 °C. The construct was diluted (1‑10, in 25 mM Tris buffer pH 7.5) loaded on anion exchange chromatography (Q Sepharose) and eluted protein loaded on a HiLoad 26/60 Superdex 75 column (GE Healthcare) using phosphate buffer (25 mM phosphate buffer pH 7.2, 300 mM NaCl). Data for the tetramerization (T) and core (C) domains of p53 were measured using the $^{13}C^{15}N$ labelled variant. The DNA used is the p53 response element (MW 1594 Da, 5’ CGCGGACATGTCCGGACATGTCCCGC’3).

Protein sequences

T (325-355) EYFTLQIRGRERFEMFRELNEALELKDAQAG

C (94-312) SSVPSQKTYQGSYGFRLGFLHSGTAKSVTCTYSPALNKLFCQLAKTCPVQLWVDSTP
PPGTRVRAMAIYKQSQHMTEVVRRCPHHERCSDSDGLAPPQHLIRVEGNLRAEYLDDR
NTFRHSVVVPYEPPEVGSDCTTIHYNYMCYSSCMGGMNRRPILTIITLEDSSGNLLGRDS
FEVRVCACPGRDRRTEEENLRKKGEPHHELPPGSTKRALPNNT

CT (94-360) SSVPSQKTYQGSYGFRLGFLHSGTAKSVTCTYSPALNKLFCQLAKTCPVQLWVDSTP
PPGTRVRAMAIYKQSQHMTEVVRRCPHHERCSDSDGLAPPQHLIRVEGNLRAEYLDDR
NTFRHSVVVPYEPPEVGSDCTTIHYNYMCYSSCMGGMNRRPILTIITLEDSSGNLLGRDS
FEVRVCACPGRDRRTEEENLRKKGEPHHELPPGSTKRALPNNTSSSPQPKKKPLDGEY
FTLQIRGRERFEMFRELNEALELKDAQAGKEPG

FL (1-393) MEEPQSDPSVEPPLSQETFSDLWKLLPENNVLSPLPSQAMDDLMLSPDDIEQWFTE
DPGPDEAPRMPEAAPPVAPAPAAPTPAAPAPAPSWPLSSSVPSQKTYQGSYGFRLGFL
HSGTAKSVTCTYSPALNKLFCQLAKTCPVQLWVDSTPPPGTRVRAMAIYKQSQHMTEV
VRRCPHHERCSDSDGLAPPQHLIRVEGNLRAEYLDDRNTFRHSVVVPYEPPEVGSDCT
TIHYNYMCYSSCMGGMNRRPILTIITLEDSSGNLLGRDSFEVRVCACPGRDRRTEEENLR
KKGEPHHELPPGSTKRALPNNTSSSPQPKKKPLDGEYFTLQIRGRERFEMFRELNEALE
LKDAQAGKEPGGSRAHSSHLKSKKGQSTSRHKKLMFKTEGPDSD

Mass Spectrometry Settings

Typical settings for the MS for the p53 constructs were: capillary voltage, 1.0-1.5 kV; sample cone 25 V; cone gas, off; trap collision voltage, 5 V; IMS drift voltage, 50-150 V; ion transfer stage pressure, 4-5 mbar; trap pressure, 8×10^{-2} mbar; ion mobility gas, He; ion mobility cell pressure, 3.2 mBar; time-of-flight analyzer pressure, $2\text{-}3\times10^{-6}$ mbar.

Experimental CCS determination

During the measurements ion mobility as well as mass spectrometric information is acquired simultaneously. Fig. S1a shows an example of a typical 2D drift time t_D vs. *m/z* contour plot, as obtained for the $(CT)_4$DNA complex at a fixed drift voltage of 50V. From this plot the arrival time distribution of each charge state is extracted (Fig. S1b) and fitted to a Gaussian distribution using python scripts developed in-house. The mobility K of each charge state is determined by plotting the drift time t_D vs. the inverse drift voltage $1/V$ and subsequent linear regression.[4] Finally, the collision cross section (CCS) Ω was calculated from the mobility K using the Mason-Schamp equation

$$\Omega = \frac{3e}{16N}\sqrt{\frac{2\pi}{\mu k_B T}}\frac{1}{K}$$

where N is the drift gas number density, μ the reduced mass of the ion and drift gas, k_B the Bolzmann constant and T the temperature. All reported CCSs are averages of at least three independent measurements.

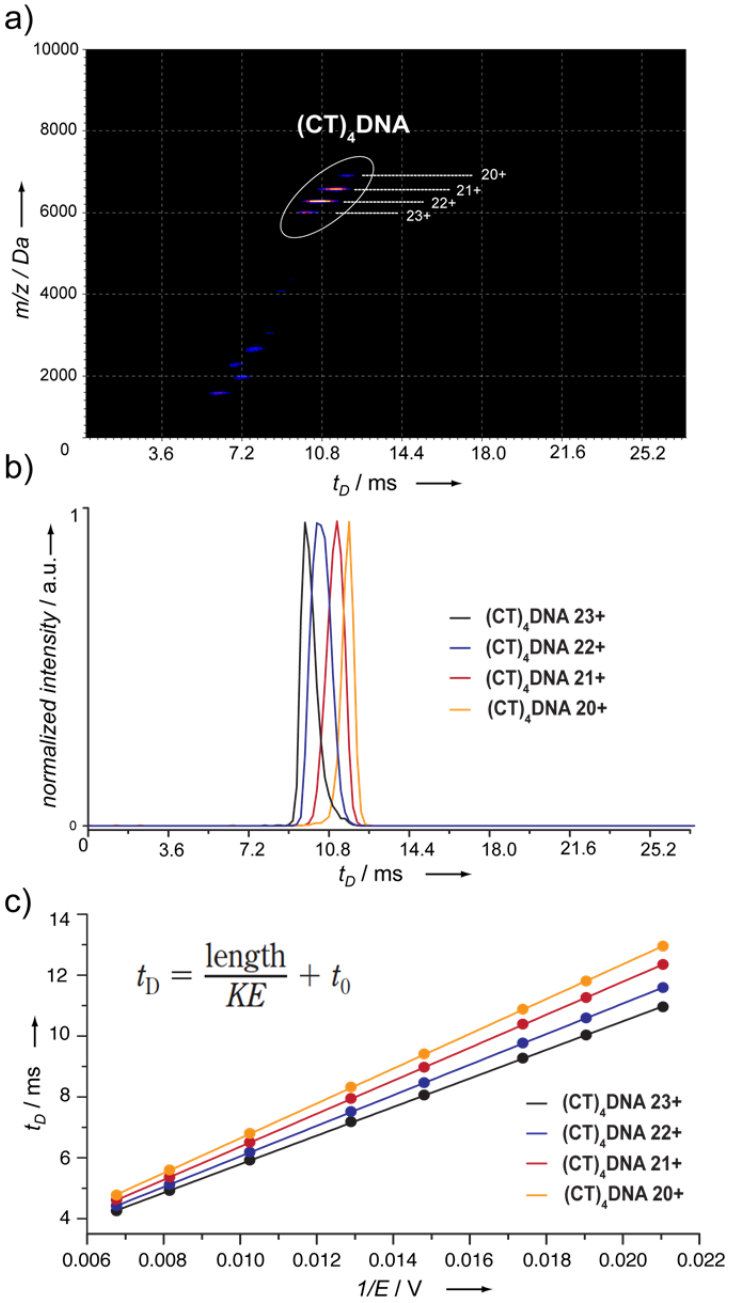

Figure S1. IM-MS data extraction and collision cross section (CCS) determination. a) Typical IM-MS contour plot of $(CT)_4$DNA which contains both drift time t_D and m/z information. Peaks corresponding to intact $(CT)_4$DNA with charge states 20 to 23+ are circled. b) Arrival time distributions extracted for the individual charge states of $(CT)_4$DNA. For each ATD the drift time is determined by Gaussian fitting. c) Plot of the inverse drift voltage vs. drift time t_D for 20 to 23+ $(CT)_4$DNA ions. After linear regression the mobility *K* was calculated from the slope of the fitted line. Collision cross sections were determined from K using the above-mentioned Mason-Schamp equation.

Calculation of theoretical CCSs

Theoretical CCSs were calculated using the PA[5] and EHSS[6] algorithms implemented in an upgraded version[7] of the widely used MOBCAL software.[5b] The following coordinates were used: $(T)_4$, *1pes*; C, *2fej*; $(C)_4$DNA, *2ac0*. The p53 segments for which no high-resolution structural information is available ($(CT)_4$, $(CT)_4$DNA, FL) were represented using coordinates from proposed models derived from a combination of X-ray, NMR, EM and SAXS data.[8]

Charge reduction

A series of previous studies showed that a controlled reduction of the charge state can stabilize gas phase protein complexes without noticeably affecting their folding behaviour in solution.[9] In order to test whether a reduced charge state has an effect on the structural collapse observed here for p53, the ammonium acetate (AA) buffer was replaced by triethylammonium acetate (TEAA). In general, this led to a reduction of the charge state (see Fig S2), in case of $(C)_4$DNA, $(CT)_4$DNA and $(FL)_4$, however, the obtained nanoelectrospray was not stable enough to enable direct CCS measurements.
For $(T)_4$, C and (CT)4 on the other hand, CCS were measured (see Table S1). The charge state turned out to have very little effect on the CCS, supporting the hypothesis of a spontaneous collapse.

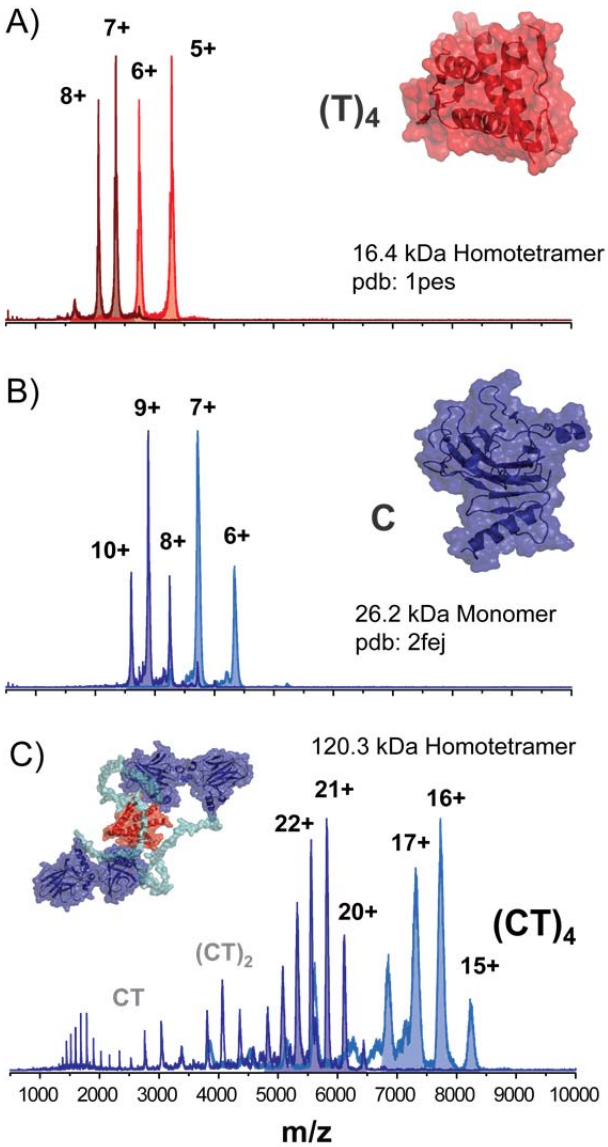

Figure S2. Representative mass spectra of charge reduced p53 constructs. (A) Spectra of the tetramerization domain tetramer $(T)_4$ sprayed from AA (dark red) and charge reducing TEAA buffer (red). (B) Spectra of the core domain monomer C sprayed from AA (purple) and TEAA buffer (blue). (C) Spectra of the core-linker-tet tetramer $(CT)_4$ sprayed from AA (purple) and charge reducing TEAA buffer (blue)

Table S1. Experimental and theoretical CCSs for p53 constructs and their DNA complexes. Values obtained form triethylammonium acetate buffered solution are highlighted in grey. (*) $^{13}C/^{15}N$ labelled variant, (**) MW_{DNA} 15946 Da, for details see SI.

Complex	Res.	MW [Da]	pdb	z	CCS exp. [Å^2]	CCS calc. [Å^2] PA	EHSS
$(T)_4$	325 - 355	16400*	1pes	8	1453	1457	1862
				7	1397		
				6	1376		
				5	1353		
C	94 - 312	26150*	2fej	10	1746	1944	2463
				9	1851		
				8	1907		
				7	1803		
				6	1840		
$(C)_4$DNA	94 - 312 + DNA**	120100*	2ac0	20	5327	5397	7098
				19	5392		
				18	5434		
$(CT)_4$	94 - 360	120300	ref 7	22	5449	8184	10319
				21	5536		
				20	5554		
				19	5521		
				16	5668		
				15	5709		
$(CT)_4$DNA	94 - 360 + DNA**	137500	ref 7	23	6012	8022	10352
				22	5978		
				21	6092		
				20	6105		
$(FL)_4$	1 - 393	176000	ref 7	29	7160	15942	19288
				28	7145		
				27	7139		
				26	7107		

Molecular dynamics simulations

MD simulations were performed in double floating-point precision using the GROMACS 4.5.3[10] software. $(CT)_4$DNA and $(CT)_4$ complexes were simulated in a solvent-free system using the AMBER99 forcefield[11] based on previously published procedures[12] to simulate proteins *in vacuo*. Briefly, a steepest descent energy minimisation was performed, followed by 1 ns vacuum simulation at constant temperature (300 K), with randomly generated initial velocities. Neither periodicity nor cut-offs were employed in the calculations. Energy conservation was achieved using a 1 fs integration step, and constraining bonds to hydrogen with the LINCS algorithm.[13] Three trajectories, for each complex were analysed with respect to their RMSD from the initial structure. In order to identify the most populated conformations at different stages in trajectories, we made use of an RMSD-clustering algorithm.[14]

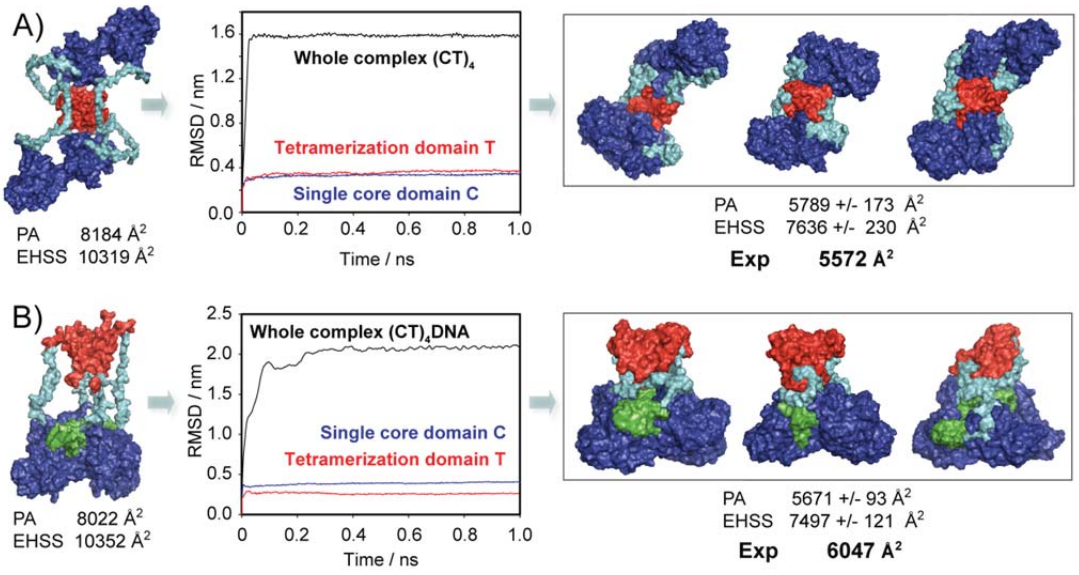

Figure S3. Molecular dynamics simulations of selected p53 constructs. MD simulations (1 ns) were carried out for (A) the $(CT)_4$ and (B) the $(CT)_4$DNA complexes *in vacuo*, at 300 K. Based on the $(CT)_4$ and $(CT)_4$DNA starting structures derived from previous experimental studies[8] (left) an RMSD-based clustering analysis was carried out, in order to identify the most populated conformations within each trajectory. Representative structures for three replica trajectories are shown in each box. Their average calculated CCSs agree well with the experimental values measured by IM-MS.

Charge state prediction

Table S2. Charge states predicted from surface area estimates. The solvent accessible surface area (SASA) was calculated from coordinate files using the MSMS program.[15] The average charge state (Z_{av}) predicted using empirical relationship $\ln(Z_{av}) = 0.59\ln(SASA)-3.12$.[16] Experimental charge state averages were calculated by Gaussian fitting of the charge state envelope after manual background subtraction and conversion into *z*-space.

Complex	Surface area (Å²)	Predicted Z_{av}	Experimental Z_{av}
$(T)_4$	8872	9.5	7.1
C	15065	13.0	9.0
$(C)_4$DNA	39909	23.2	19.1
$(CT)_4$	62106	30.2	21.4
$(CT)_4$DNA	65276	31.1	21.7
$(FL)_4$	121133	44.8	28.8
$(CT)_4$ (MD)	36203	21.9	21.4
$(CT)_4$DNA (MD)	35568	21.6	21.7

References

[1] a) A. C. Joerger, M. D. Allen, A. R. Fersht, *J. Biol. Chem.* **2004**, *279*, 1291-1296; b) P. V. Nikolova, J. Henckel, D. P. Lane, A. R. Fersht, *Proc. Natl. Acad. Sci. U S A* **1998**, *95*, 14675-14680.

[2] E. Arbely, E. Natan, T. Brandt, M. D. Allen, D. B. Veprintsev, C. V. Robinson, J. W. Chin, A. C. Joerger, A. R. Fersht, *Proc. Natl. Acad. Sci. U S A* **2011**, *108*, 8251-8256.

[3] E. Natan, A. C. Joerger, *J. Mol. Biol*. **2012,** *415*, 503-13

[4] M. F. Bush, Z. Hall, K. Giles, J. Hoyes, C. V. Robinson, B. T. Ruotolo, *Anal. Chem.* **2010**, *82*, 9557–9565.

[5] a) N. G. Gotts, G. von Helden, M. T. Bowers, *International Journal of Mass Spectrometry and Ion Processes* **1995**, *149-150*, 217-229; b) M. F. Mesleh, J. M. Hunter, A. A. Shvartsburg, G. C. Schatz, M. F. Jarrold, *J. Phys. Chem.* **1996**, *100*, 16082-16086.

[6] A. A. Shvartsburg, M. F. Jarrold, *Chemical Physics Letters* **1996**, *261*, 86-91.

[7] A. A. Shvartsburg, S. V. Mashkevich, E. S. Baker, R. D. Smith, *J. Phys. Chem. A* **2007**, *111*, 2002-2010.

[8] H. Tidow, R. Melero, E. Mylonas, S. M. V. Freund, J. G. Grossmann, J. M. Carazo, D. I. Svergun, M. Valle, A. R. Fersht, *Proc. Natl. Acad. Sci. U S A* **2007**, *104*, 12324-12329.

[9] a) D. Lemaire, G. Marie, L. Serani, O. Laprevote, *Anal. Chem.* **2001**, *73*, 1699-1706; b) K. Pagel, S.-J. Hyung, B. T. Ruotolo, C. V. Robinson, *Anal. Chem.* **2010**, *82*, 5363-5372.

[10] D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, H. J. C. Berendsen, *J. Comput. Chem.* **2005**, *26*, 1701-1718.

[11] J. Wang, P. Cieplak, P. A. Kollman, *J. Comput. Chem.* **2000**, *21*, 1049-1074.

[12] a) E. G. Marklund, D. S. D. Larsson, D. van der Spoel, A. Patriksson, C. Caleman, *Physical Chemistry Chemical Physics* **2009**, *11*; b) Z. Hall, A. Politis, M. F. Bush, L. J. Smith, C. V. Robinson, *J. Am. Chem. Soc.* **2012**.

[13] B. Hess, H. Bekker, H. J. C. Berendsen, J. G. E. M. Fraaije, *J. Comput. Chem.* **1997**, *18*, 1463-1472.

[14] X. Daura, K. Gademann, B. Jaun, D. Seebach, W. F. van Gunsteren, A. E. Mark, *Angew. Chem. Int. Ed.* **1999**, *38*, 236-240.

[15] M. F. Sanner, A. J. Olson, J.-C. Spehner, *Biopolymers* **1996**, *38*, 305-320.

[16] Z. Hall, C. V. Robinson, *J. Am. Soc. Mass. Spectrom.* **2012**, in press.