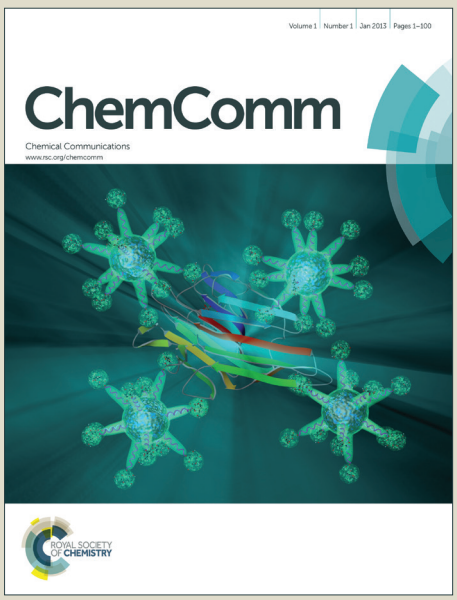
ChemComm

Accepted Manuscript



ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/chemcomm

Journal Name

RSCPublishing

COMMUNICATION

A role for a disordered domain in regulating protein oligomerization and stability

Cite this: DOI: 10.1039/x0xx00000x

Received 00th January 2012,
Accepted 00th January 2012

DOI: 10.1039/x0xx00000x

www.rsc.org/

Ofrah Faust[a], Lavi Bigman[a] and Assaf Friedler[a*]

Intrinsically disordered proteins (IDPs) or regions (IDRs) in proteins hold many functions but their biological roles are still not fully understood. Here we describe a new role for such regions. Using the HIV-1 Rev protein, we show that disordered domains have a role in maintaining the correct oligomeric state and the thermodynamic stability of proteins.

Intrinsically disordered proteins (IDPs) lack a stable tertiary structure and their active state is an ensemble of interchanging flexible conformations. IDPs have low overall hydrophobicity and a large net charge, preventing formation of a hydrophobic core and thus proper folding.[1] Intrinsically disordered regions (IDRs) appear in many types of proteins and can possess many functions such as DNA binding, protein binding, intramolecular binding and more[2,3]. About one third of the eukaryotic proteins are highly disordered and more than 50% contain IDRs.[4] IDPs and IDRs often form interactions with several partners as a result of their ability to gain different structures upon binding different targets[4,5].

IDPs are also abundant in viruses, probably due to their need to be genetically compact and to have higher adaptability to mutations.[6] HIV-1 has a number of IDPs or proteins with IDRs: Tat, a transactivator of viral transcription, is intrinsically disordered[7]; Vpr, an accessory protein with multiple activities is disordered at neutral pH[8]; Vif, an accessory protein that neutralizes the anti-viral cellular defense mechanism has a disordered C-terminal domain that may gain structure upon binding[9]; Nef, an accessory protein, has disordered regions that undergo conformational changes upon binding to their ligands[10].

The HIV-1 Rev protein is a small 116 residues, 13 kDa protein that shuttles the unspliced or partially spliced viral RNA from the host cell nucleus to the cytoplasm[11]. This enables packing of the viral RNA into the emerging virions, resulting in continuous infectivity. A Rev multimer shuttles the RNA out of the nucleus[12] by binding specifically to the Rev Response Element (RRE). The exact number of Rev monomers that bind the RRE in cells is not known. Up to eight or even ten Rev monomers can bind the RRE simultaneously[13,14]. The binding is cooperative[14–17]: One Rev monomer binds the high affinity site on stem-loop IIB on the RRE and additional monomers join through protein-protein and protein-RNA interactions, one at a time.

Rev consists of a structured N-terminal domain (Rev NTD, residues 1-65) and a disordered C-terminal domain (Rev CTD, residues 66-116) (Fig. 1). The structure of the Rev NTD as part of a Rev-Fab complex was solved using X-ray crystallography[18] and matched former predictions and experiments[19,20]: A helix-loop-helix motif resides between residues 9-65, while the rest (~50%) of the protein remains disordered. The structure showed a Rev dimer, but it was also found that Rev forms tetramers in solution[21]. Using point mutations, four residues - I19, L22, I52 and I59 - have been proposed to contribute to the intramolecular interface, stabilizing the helices[22]. The arginine rich motif (ARM), which holds the specific RRE-RNA binding site and the nuclear localization signal (NLS), is located between residues 34-50 (Fig. 1a). The leucine-rich nuclear export signal (NES) is found in the disordered part of the protein, between residues 74-83. It is not known what the role of the rest of the disordered domain (residues 65-73, 84-116) is.

Here we present a structural and functional analysis of the Rev CTD. We found that the Rev CTD contributes to the thermodynamic stability of the protein, increases its T_m and is essential in forming its proper oligomeric states. Both correct oligomerization and stability are essential for the proper activity of proteins in the cell. Our results present a new role for IDRs.

We ran disorder predictions of the full length Rev sequence using several disorder prediction servers (Fig. 1b). The CTD was predicted to be mostly disordered. This is in agreement with the results reported for the crystal structure of Rev (Fig. 1c), in which the CTD showed no electron density[18], typical of a disordered domain. To study the role of the Rev CTD, we expressed and purified the two Rev domains: Rev 1–65 (Rev NTD) and Rev 66–116 (Rev CTD). Both fragments were labeled by a tag consisting of a His tag, a Lipoyl domain and a TEV cleavage site, to assist in the expression, purification and solubilization of the protein.

The domains were expressed and purified as described in the ESI. MALDI-TOF MS confirmed the identity of both fragments. The full-length Rev was expressed and purified as we described previously[31]. Using CD, we analyzed the secondary structures of the Rev domains. We found that both Rev NTD and full-length Rev are structured: The CD spectra of both proteins showed minima at 222 nm and at 208 nm, characteristic of an alpha-helical structure (Fig. 2). The Rev CTD displayed a CD spectrum characteristic of disordered proteins, with one minimum around 200 nm. To our

knowledge, this is the first direct experimental evidence of the disordered nature of this domain.

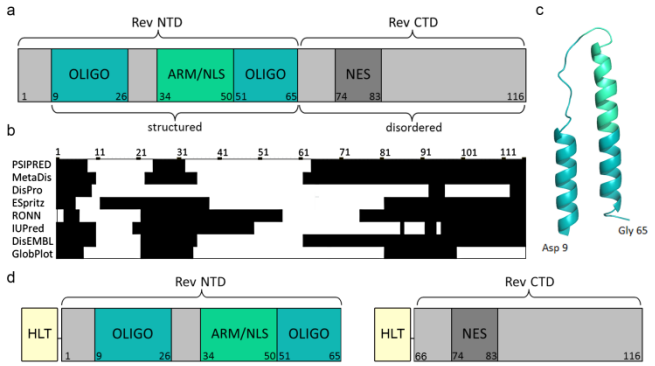


Figure 1. The domains of Rev. A. The N-terminal domain (Rev NTD) resides between residues 1-65 and includes the oligomerization domains, the arginine rich motif (ARM) and the nuclear localization signal (NLS). The C-terminal domain (Rev CTD) includes residues 66-116 that contain the nuclear export signal (NES) and is intrinsically disordered; B. Disorder predictions of Rev using seven different servers[23–30]. Most servers predict a highly disordered CTD; C. The structure of Rev NTD (PDB: 2X7L[18]). It is mostly structured, consisting of a helix-loop-helix. D. Bar representation of the Rev fragments used in this research.

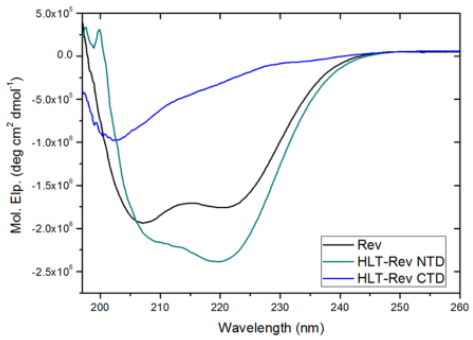


Figure 2. The Rev CTD is disordered while the Rev NTD is structured. Shown are CD spectra of Rev (black), HLT-Rev NTD (cyan) and HLT-Rev CTD (dark blue). Rev and HLT-Rev NTD show an alpha-helical structure, while the CTD is disordered. The role of the CTD in regulating Rev oligomerization.

To study the role of Rev CTD in the oligomerization of the protein, we used analytical size-exclusion chromatography (SEC). Rev and Rev NTD were run on an analytical Superose 12 column (GE Healthcare). We discovered a significant difference in their elution profile (Fig. 3a). Both proteins eluted in a single major peak and a second smaller peak that partly overlapped the major peak. Elution in two peaks indicates existence of two populations corresponding to two oligomeric states, but here these were different for the two proteins. Full-length Rev eluted primarily at 3.1 ml and had a shoulder at a lower elution volume, indicating the existence of several oligomeric populations as observed before[32]. The dominant population was in the lower oligomeric state, which was impossible to quantify due to the presence of the disordered domain. Rev NTD eluted primarily at 2.3 ml and showed a small peak at a higher elution volume of 3.1 ml. This means that Rev NTD exists mainly in a high oligomeric state with a smaller population in a lower state, in contrast to the full-length Rev populations. The results show that presence of the CTD in the full-length Rev restricted the oligomers to low order ones and prevented aggregation of the protein.

To further characterize the oligomeric states of Rev and Rev NTD, we used cross-linking with bis-(sulfosuccinimidyl) suberate (BS3) followed by SDS-PAGE. Both proteins formed oligomers, as expected. The full length Rev displayed several low order oligomeric states (e.g. dimers, trimers and tetramers), seen as distinct bands in the gel (Fig. 3b). Very little high order oligomers were observed. Rev NTD existed in oligomers of very high order that hardly entered the gel, as well as in a small amount of monomers and dimers. This is in agreement with the SEC results: Rev NTD alone formed high order oligomers, and only the presence of the disordered CTD restricted the number of oligomers to small numbers preventing formation of large aggregates.

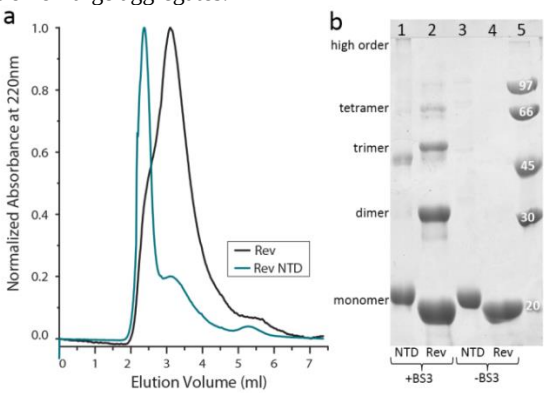


Figure 3. The role of the CTD in regulating Rev oligomerization. A. The elution profiles of Rev (black) and of Rev NTD (cyan) in size-exclusion chromatography. Both proteins form different oligomeric states; B. With no cross linker, the proteins are denatured and run as one band in the gel (lanes 3, 4). With the cross linker, Rev forms oligomers of several sizes (lane 2) while Rev NTD (lane 1) forms higher order oligomers or aggregates. Marker molecular weights are shown in lane 5, in kDa.

Analytical ultracentrifugation (AUC) sedimentation velocity experiments also revealed that the full length protein existed in one major population, with a sedimentation coefficient of ~2S (Fig. 4a). Under the same conditions, Rev NTD showed a variety of states, the lowest of which was around 2S and the others between 4S and 5S (Fig. 4b). Higher sedimentation coefficients indicate a higher oligomeric state. This further emphasizes the role of Rev CTD in regulating protein oligomerization, maintaining the lower state and preventing aggregation. The HLT tag itself did not induce oligomerization, as can be seen in figure S1. The tag alone did not oligomerize, and neither did HLT-Rev CTD.

To test the effect of the Rev CTD on the stability of the protein, we used temperature dependent CD and obtained a melting curve for full length Rev and Rev NTD (Fig. 5). While the T_m for full length Rev was 38.2±0.1 °C, the T_m for Rev NTD alone was 34.1±0.1 °C. This indicates that the Rev CTD has a role in thermodynamically stabilizing the full length protein.Disordered domains were shown before to regulate proteins activity by an intramolecular interaction with the structured domains[33,34]. We used several methods to test whether this mode of regulation of intramolecular binding between the two domains is valid also for Rev (figure S2). However, the Rev CTD did not interact directly with the NTD (see supporting information for details).

The results presented herein demonstrate that the Rev CTD regulates the stability and the correct oligomeric state of Rev. Maintaining the correct oligomeric state is crucial for the proper function of proteins. In our specific model system, the correct oligomeric state of the Rev protein results in HIV-1 infectivity. While the structured NTD is the domain responsible for the actual oligomerization, it is the presence of the CTD that restricts the number of oligomers to a small number and prevents uncontrolled aggregation of the protein. A Rev protein lacking the CTD has a T_m under 37 °C, which means it is mostly denatured and not functional in the body.

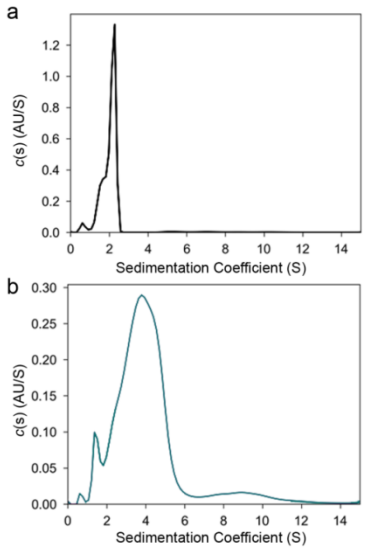


Figure 4. The presence of the CTD regulates Rev oligomerization: AUC studies. A. Rev has a sedimentation coefficient of ~2S. B. HLT-Rev NTD showed a wider distribution, with a sedimentation coefficient of ~4S.

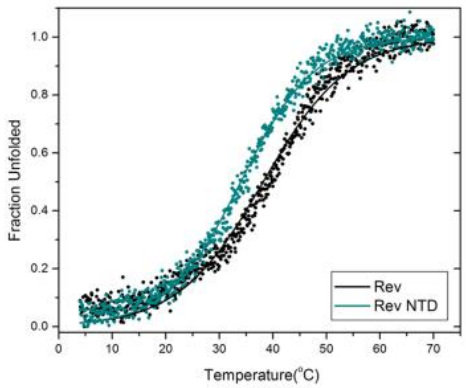


Figure 5. The presence of the CTD increases the thermodynamic stability of Rev. Shown are melting curves for Rev (black) and HLT-Rev NTD (cyan). Data (dots) were fit (solid line) to a sigmoidal model. The T_m of the full length protein is 38.2°±0.1C while the T_m of Rev NTD is 34.1±0.1°C.

The presence of the disordered CTD provides Rev the extra thermodynamic stability that enables it to be active at body temperature, but the protein is still only marginally stable. Many proteins have evolved to be only marginally stable, functioning only at a temperature that suits their environment. This enables tight regulation and control of the activity of these proteins. We found that Rev is such a protein, which is in agreement with its regulatory role in the HIV life cycle. The low melting temperature and marginal stability of Rev can also be part of the reason why Rev is a very difficult to express protein with a high tendency to aggregate and to form fibrils[35]. Indeed, an efficient expression protocol for Rev was developed by us only recently[31], and its crystal structure was solved also only very recently with an engineered Fab antibody[18,36]. We experienced that the expression and purification of the structured NTD was more difficult than that of the full-length protein, proving the need for the presence of the CTD for stabilizing the protein.

Disordered domains were shown to regulate the activity of structured protein domains by intramolecular binding as we have shown for the ASPP2 protein [33,34] and were also shown to solubilize[37] and to stabilize[38] proteins. Many of the regulatory roles carried out by IDRs are carried out through protein-protein interactions and direct binding. However, no direct intramolecular binding was detected here between the two Rev domains, indicating that the mode of regulation here is different than what was shown before. The presence of the CTD is sufficient for stabilizing the full length protein and regulating its oligomerization.Rev binds the RRE in a well characterized cooperative mechanism. We suggest that the regulation by the CTD is important for keeping Rev in a functional state that allows this mechanism to take place. A large Rev oligomer cannot bind the RRE, and in such a case the shuttling of the RNA would be inhibited. The regulation by the CTD keeps Rev in the low oligomeric state that enables the sequential binding to the RRE and consequently the nuclear export of the viral RNA. The oligomerization of Rev must be well regulated for efficient protein activity and virus infectivity. The regulation by the CTD without binding the NTD could be a mechanism enabling the high mutation rate found in HIV. Since there are no specific binding residues, mutations in Rev can be tolerated without loss of activity.

The Rev IDR is located in its C-terminus. This location enables it to be highly flexible and its movement is not confined by flanking structured domains. This way the hindrance is maximized, enabling the IDR to potentially bind many regions in the structured NTD. Many IDRs are located at the termini of proteins[2,39,40], and their role is not always understood. We believe that they could also act as shields and regulate the quaternary structure of the protein.

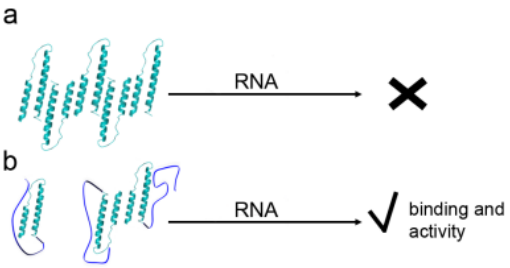


Figure 6. A Model for how Rev CTD regulates the protein oligomerization and stability. A. In absence of the Rev CTD, all oligomerization domains are exposed, enabling the formation of high-order oligomers that are unable to bind the RRE. Here Rev is inactive. B. The disordered CTD acts as a shield. The actual presence of the CTD limits the oligomerization state to the lower order oligomers and enables the protein to be in its stable and active form. Rev can bind the RRE, which results in viral infectivity.

In summary, we suggest a model of regulation in which the actual presence of the disordered CTD limits the number of monomers that can join the Rev multimer. This can be done by steric hindrance of the disordered CTD or by transient interactions, forming a dynamic complex. (Fig. 6). This hindrance prevents addition of monomers and maintains the correct low oligomeric form, in which the protein is most stable and active, by shielding the oligomerization-mediating residues and preventing additional monomers from joining**.** Our findings show a new role for the Rev IDR in regulating protein oligomerization and stability. Rev has an IDR that decreases the number of monomers in the oligomer, prevents aggregation and raises the T_m of the protein. This is a new role for IDRs, which may be true for many IDRs in other biological systems.

Notes and references

[a] Institute of Chemistry, The Hebrew University of Jerusalem, The Edmond J. Safra Campus, Givat Ram 91904, Jerusalem, Israel. AF was supported by a starting grant from the European Research Council under the European Community's Seventh Framework Programme (FP7/2007-2013) / ERC Grant agreement n° 203413 and by the Minerva Center for

Bio-hybrid complex systems. Electronic Supplementary Information (ESI) available: experimental methods and binding experiments.

1. V. N. Uversky and a K. Dunker, *Biochim. Biophys. Acta*, 2010, **1804**, 1231–64.

2. V. N. Uversky, *FEBS Lett.*, 2013, **587**, 1891–901.

3. M. M. Babu, R. W. Kriwacki, and R. V Pappu, *Science*, 2012, **337**, 1460–1.

4. M. S. Cortese, V. N. Uversky, and a K. Dunker, *Prog. Biophys. Mol. Biol.*, 2008, **98**, 85–106.

5. H. Amartely, A. David, M. Lebendiker, H. Benyamini, S. Izraeli, and A. Friedler, *Chem. Commun. (Camb).*, 2013.

6. V. N. Uversky and S. Longhi, *Flexible Viruses Structural Disorder in Viral Proteins*, Wiley, 2012.

7. S. Shojania and J. D. O'Neil, *J. Biol. Chem.*, 2006, **281**, 8347–56.

8. P. Henklein, K. Bruns, M. P. Sherman, U. Tessmer, K. Licha, J. Kopp, C. M. de Noronha, W. C. Greene, V. Wray, and U. Schubert, *J. Biol. Chem.*, 2000, **275**, 32016–26.

9. T. H. Reingewertz, H. Benyamini, M. Lebendiker, D. E. Shalev, and A. Friedler, *Protein Eng. Des. Sel.*, 2009, **22**, 281–7.

10. S. a Leavitt, A. SchOn, J. C. Klein, U. Manjappara, I. M. Chaiken, and E. Freire, *Curr. Protein Pept. Sci.*, 2004, **5**, 1–8.

11. V. W. Pollard and M. H. Malim, *Annu Rev Microbiol*, 1998, **52**, 491–532.

12. S. P. Edgcomb, A. Aschrafi, E. Kompfner, J. R. Williamson, L. Gerace, and M. Hennig, *Protein Sci*, 2008, **17**, 420–430.

13. D. A. Mann, I. Mikaelian, R. W. Zemmel, S. M. Green, A. D. Lowe, T. Kimura, M. Singh, P. J. Butler, M. J. Gait, and J. Karn, *J Mol Biol*, 1994, **241**, 193–207.

14. T. J. Daly, R. C. Doten, P. Rennert, M. Auer, H. Jaksche, A. Donner, G. Fisk, and J. R. Rusche, *Biochemistry*, 1993, **32**, 10497–10505.

15. M. H. Malim and B. R. Cullen, *Cell*, 1991, **65**, 241–248.

16. K. S. Cook, G. J. Fisk, J. Hauber, N. Usman, T. J. Daly, and J. R. Rusche, *Nucleic Acids Res*, 1991, **19**, 1577–1583.

17. S. J. Pond, W. K. Ridgeway, R. Robertson, J. Wang, and D. P. Millar, *Proc Natl Acad Sci U S A*, 2009, **106**, 1404–1408.

18. M. A. DiMattia, N. R. Watts, S. J. Stahl, C. Rader, P. T. Wingfield, D. I. Stuart, A. C. Steven, and J. M. Grimes, *Proc Natl Acad Sci U S A*, 2010, **107**, 5810–5814.

19. F. J. Blanco, S. Hess, L. K. Pannell, N. W. Rizzo, and R. Tycko, *J Mol Biol*, 2001, **313**, 845–859.

20. J. L. Battiste, H. Mao, N. S. Rao, R. Tan, D. R. Muhandiram, L. E. Kay, a D. Frankel, and J. R. Williamson, *Science*, 1996, **273**, 1547–51.

21. M. L. Zapp, T. J. Hope, T. G. Parslow, and M. R. Green, *Proc. Natl. Acad. Sci.*, 1991, **88**, 7734–7738.

22. C. Jain and J. G. Belasco, *Mol Cell*, 2001, **7**, 603–614.

23. Z. Dosztányi, V. Csizmok, P. Tompa, and I. Simon, *Bioinforma.* , 2005, **21** , 3433–3434.

24. L. J. McGuffin, K. Bryson, and D. T. Jones, *Bioinforma.* , 2000, **16** , 404–405.

25. I. Walsh, A. J. M. Martin, T. Di Domenico, and S. C. E. Tosatto, *Bioinforma.* , 2012, **28** , 503–509.

26. Z. R. Yang, R. Thomson, P. McNeil, and R. M. Esnouf, *Bioinformatics*, 2005, **21**, 3369–76.

27. L. P. Kozlowski and J. M. Bujnicki, *BMC Bioinformatics*, 2012, **13**, 111.

28. R. Thomson and R. Esnouf, in *Intelligent Data Engineering and Automated Learning – IDEAL 2004 SE - 16*, eds. Z. Yang, H. Yin, and R. Everson, Springer Berlin Heidelberg, 2004, vol. 3177, pp. 108–116.

29. Z. Dosztányi, V. Csizmók, P. Tompa, and I. Simon, *J. Mol. Biol.*, 2005, **347**, 827–839.

30. R. Linding, R. B. Russell, V. Neduva, and T. J. Gibson, *Nucleic Acids Res.* , 2003, **31** , 3701–3708.

31. P. Siman, O. Blatt, T. Moyal, T. Danieli, M. Lebendiker, H. a Lashuel, A. Friedler, and A. Brik, *ChemBioChem*, 2011, **12**, 1097–1104.

32. M. D. Daugherty, D. S. Booth, B. Jayaraman, Y. Cheng, and A. D. Frankel, *Proc. Natl. Acad. Sci. U. S. A.*, 2010, **107**, 12481–6.

33. S. Rotem, C. Katz, H. Benyamini, M. Lebendiker, D. Veprintsev, S. Rüdiger, T. Danieli, and A. Friedler, *J. Biol. Chem.*, 2008, **283**, 18990–9.

34. S. Rotem-Bamberger, C. Katz, and A. Friedler, *PLoS One*, 2013, **8**, e58470.

35. P. T. Wingfield, S. J. Stahl, M. A. Payton, S. Venkatesan, M. Misra, and A. C. Steven, *Biochemistry*, 1991, **30**, 7527–7534.

36. S. J. Stahl, N. R. Watts, C. Rader, M. A. DiMattia, R. G. Mage, I. Palmer, J. D. Kaufman, J. M. Grimes, D. I. Stuart, A. C. Steven, and P. T. Wingfield, *J. Mol. Biol.*, 2010, **397**, 697–708.

37. V. Bandaru, W. Cooper, S. S. Wallace, and S. Doublié, *Acta Crystallogr. D. Biol. Crystallogr.*, 2004, **60**, 1142–4.

38. M. L. Hegde, S. E. Tsutakawa, P. M. Hegde, L. M. F. Holthauzen, J. Li, N. Oezguen, V. J. Hilser, J. a Tainer, and S. Mitra, *J. Mol. Biol.*, 2013, **425**, 2359–71.

39. T. H. Reingewertz, D. E. Shalev, S. Sukenik, O. Blatt, S. Rotem-Bamberger, M. Lebendiker, S. Larisch, and A. Friedler, *PLoS One*, 2011, **6**, e24655.

40. M. M. Rosenberg, D. Ronen, N. Lahav, E. Nazirov, S. Ravid, and A. Friedler, *J. Biol. Chem.*, 2013, **288**, 9779–89.