

An automated workflow for the discovery and docking simulation of the protein-protein complexes using *in vivo* chemical cross-linking

Vladimir Svetlov¹ and Evgeny Nudler^{*1,2}

1 – Department of Biochemistry and Molecular Pharmacology, New York University School of Medicine, New York, NY 10016, USA; 2 – Howard Hughes Medical Institute, Chevy Chase, MD 20815, USA; * – corresponding author.

Introduction

Chemical cross(X)-link mapping assisted by mass spectrometry (XL-MS, also CXMS and CLMS) is a low-resolution hybrid method of structural biology, yielding a set of pairwise distance restraints between reactive solvent-accessible amino acids¹⁻⁹. Most commonly used X-linkers at present belong to the class of amino-reactive homo-bifunctional NHS-esters, which act as protein proximity sensors, connecting predominantly Lys residues (as well as far less common and informative N-terminal amines)¹⁰⁻¹⁴. Main structural application of XL-MS to date is limited to the generation of distance restraints assisting in interpretation of the cryo-EM data^{3,5,15-17}. This approach usually entails treating the same sample used in cryo-EM experiments (purified protein-protein complex (PPC) often assembled from over-expressed components) *in vitro* with chemical X-linkers prior to the entry into the XL-MS pipeline. In that XL-MS relies on an orthogonal information regarding PPC composition as do other structural methods, thereby limiting its exploratory potential.

Chemical X-linking has been used in proteomics and interactomics applications long before the advent of XL-MS. Most commonly *in vivo* interactomics approach utilizes small, broadly specific, cell-permeable X-linkers such as formaldehyde, administered to the culture/or suspension of cells (live tissue applications face an obvious impediment to X-linker diffusion due to multi-layering of membranes and cell walls, as well as intra-cellular macromolecular crowding)¹⁸⁻²¹. X-linked PPCs of interest are then purified, using affinity (including immuno-affinity) tags, and subjected to proteolysis followed by MS/MS identification of linear peptides(AP/IP-MS)²²⁻²⁴. This enumerative proteomics is made possible by the fact that X-linkers react only with solvent-accessible amino acids, leaving buried/interior peptides unmodified and thus amenable to straightforward shotgun protein ID.

Non-labile covalent X-links formed *in vivo* antagonize dissociation of the PPC during cell disruption, extraction and purification steps, thereby reducing generation of the false negative data points compared to affinity/immune-purification carried out without X-linking. However, lacking means to distinguish co-purification from specific PPC formation, it does not prevent inclusion of the false positive points into the dataset, which often includes hundreds of potential interactors for any given “bait” (affinity-tagged target protein). In contrast, XL-MS by providing objective measure of proximity and unique PPC architecture (as a non-random set of X-linked peptides) allows for an unbiased removal of the false positives from the data.

XL-MS use in *in vivo* interactomics has been rather limited. Formaldehyde, although widely used in X-linking applications, has broad reactivity towards amino acids, thereby complicating automated (rule-based) discovery²¹. In contrast, popular XL-MS cross-linkers often exhibit low cell-permeability, due to the presence of easily ionizable groups (e.g. bis(sulfosuccinimidyl)suberate (BS3)) or bulky spacers (e.g. functionalized reagents featuring affinity or clickable groups). Successful forays into XL-MS of live cells have been rather scarce, with one of the most successful platforms developed for *in vivo* XL-MS utilizes cleavable NHS-ester X-linker disuccinimidyl sulfoxide (DSSO) and LC-MSⁿ-assisted identification of the X-linked peptides²⁵. This platform offers many advantages to a highly skilled operator, but its reliance on in-house expertise and scripts certainly restricts its portability to the labs that do not have access to either. In this work we set to assemble a streamlined *in vivo* XL-MS workflow, from chemical X-linking to X-link-assisted protein-protein docking, which utilizes broadly available reagents and automated (score-driven) computational tools and thus requires minimal expert intervention.

Results

Discovery of *E. coli* NusA-RNA polymerase (RNAP) interactions by *in vivo* XL-MS.

Water soluble X-linker BS3 has been successfully used in structural interrogation of PPCs by XL-MS *in vitro* due to its preference for Lys as the reactive group, predictable fragmentation of X-linked peptides, and commercial

availability^{16,17,26-29}. Its cell-permeable isostere disuccinimidyl suberate (DSS) was routinely deployed in AP/IP-MS proteomics experiments, as well as in *in vitro* XL-MS applications³⁰⁻³⁵. To extend its use to *in vivo* XL-MS we have treated an exponentially growing culture of *E. coli* RL721 (*rpoC::His6*) with 2 mM DSS for 30 min at 37°C, RNAP polymerase and the associated proteins were purified using Ni²⁺-affinity resin, and digested with trypsin for LC-MS/MS analysis. Each sample was divided in two aliquots, the first were processed for linear peptide discovery (data-dependent acquisition of ions charged 1, 2, and 3), and the second – for X-linked peptides discovery (data-dependent acquisition of ions charged 4, 5, and 6).

In order to reduce non-specific binding of proteins to the resin purification was carried out in denaturing conditions (in presence of 8M urea). The number of co-purifying proteins identified in the denatured samples was reduced compared to the same discovered in the samples purified without the addition of urea, although not entirely free from known contaminants and other unrelated proteins (data not shown). As expected from the His₆-tagged RpoC-baited samples, the list of co-purified proteins included RpoC, other RNAP subunits (RpoB, RpoA, RpoZ), and RNAP-associated initiation and elongation factors (RpoD, RpoS, NusA, NusG, GreA, GreB, DksA, etc), as well as unrelated proteins of *E. coli* and human origins (VS and EN, manuscript in preparation). *fasta* files of all corresponding proteins were extracted from the Uniprot and pooled together to form the search database (combined *fasta* file) for XL-MS. The reduction of the search space to the proteins actually present in the sample lowers the computational cost (compared to the search of the entire proteome), but lacking any other filtering of the input, it prevents biasing the discovery process.

XL-MS results obtained by pLink1 were filtered during the search and recording of the output by the false discovery rate (FDR<5%) and e-value (≤ 0.001)³⁶. As expected from the RpoC-baited AP-XL-MS, discovered X-links originated predominantly in one of the RNAP subunits (VS and EN, manuscript in preparation). No X-links between RpoA/B/C/Z proteins and known contaminants (e.g. metal-binding metabolic enzymes)³⁷ were detected in the sample, indicating that this approach didn't introduce any false positives into the dataset. Consistent with orthogonal interactomics data, a number X-links were detected between RNAP subunits and potential interactors, such as initiation and elongation factors RpoD, Rho, DksA, NusG, and NusA. Efficient discovery of these X-links in a crude, single-tag AP is most likely due to the high abundance of these factors, comparable with that of RpoC (7164 molecules/cell in LB): RpoD – 1657, Rho – 5934, DksA – 19594, NusG – 6938, and NusA – 10025 molecules/cell, respectively³⁸. It bears noting that no X-links were found between RNAP subunits and highly abundant unlikely interactors, such as SucB (15565), GltA (16361), and CysK (21811, all – molecules/cell). The apparent lack of the experimental X-links between Rpo proteins and low abundance transcriptional regulators present in the sample, such as MalT (75 molecules/cell) and ExuR (94 molecules/cell), indicates that this approach is prone to generate false negatives in regard to under-represented proteins, when highly abundant bait is used. This predicament may be endemic to the highly complex RNAP interactome, which at any given time consists of hundreds of transcription regulators present at intracellular concentrations spanning 4 orders of magnitude³⁹⁻⁴¹.

For the purpose of generating an automated PPC docking model based on *in vivo* XL-MS data we selected the binary interaction between *E. coli* RNAP and the general transcription elongation regulator NusA. NusA was found in the cell in the concentration nearly stoichiometric to that of RNAP subunits, and is believed to be present in the majority, if not the entirety of the actively transcribing elongation complexes^{42,43}. NusA plays a variety of roles in transcript elongation and termination, as well as in a continually growing number of transcription-associated/-coupled cellular functions, pertaining to DNA repair and genome stability, protein traffic on DNA, etc⁴⁴⁻⁵⁴. The availability of orthogonal structural information regarding NusA-RNAP allows for a confident benchmarking of the docking model. In XL-MS experiments NusA-RNAP X-links were highly abundant and reproducible, consistent with the existence of a specific and persistent PPC.

Majority of the *in vivo* NusA-RNAP X-links exhibited clustering in the N-terminal domain (NTD) of NusA and the vicinity of the so-called beta flap domain of RNAP, with the remainder scattered around flexible regions in the C-terminal half of NusA and flexible/unstructured fragments of RpoA subunit(s) of RNAP. Hence we have selected the following set of *in vivo* X-links formed between NusA NTD and RNAP for the docking simulation: NusA3-RpoB900, NusA37-RpoB890, NusA37-RpoB909, NusA38-RpoB890, NusA38-RpoB909, NusA111-RpoB890, NusA111-RpoB909, and NusA143-RpoC50.

X-links-guided docking of NusA NTD and RNAP

There are a few protein-protein docking algorithms/servers, with HADDOCK standing out as the one that performs consistently well in the automated mode, with the starting structural files and unambiguous distance restraints as the only input from the user⁵⁵. Lacking a high quality experimental structure of the full-length NusA, we have modeled its NTD (residues 1-180) using one of the top performing automated homology modeling servers, I-

TASSER⁵⁶. Resulting high-confidence model (C-score=0.98, estimated TM-score=0.85±0.08, estimated RMSD=2.7±2.0 Å) was further refined using FG-MD server⁵⁷ to yield NusA NTD starting structure file, chain A. Since the only experimental inter-protein X-links formed by NusA NTD were limited to the RNAP two largest subunits, RpoB and RpoC, we have extracted coordinates of those subunits from experimental X-ray crystallographic model 4lk1⁵⁸. Chains corresponding to RpoB and RpoC were joined into one (chain B), numbering of the RpoB residues was left unaltered, whereas RpoC residues were re-numbered by adding 2000 to their original numbers.

Unambiguous restraints were entered as the pairwise NZ distances between residues NusA3-RpoB900, NusA37-RpoB890, NusA37-RpoB909, NusA38-RpoB890, NusA38-RpoB909, NusA111-RpoB890, NusA111-RpoB909, and NusA143-RpoC50, conservatively set at 10.0 Å as the target distance, 4.0 Å as the lower correction, and 3.0 Å as the upper correction^{59,60}. Docking was performed using the Expert Interface at the HADDOCK server, yielding 200 docking models divided between 2 clusters. Cluster 1 comprised 117 models and had higher HADDOCK score (127.0±9.0), greater buried surface area (1901.2±134.6), and lower restraints violation penalty (87.9±6.99), than cluster 2 (for the detailed comparison see Figure 1 and Table 1). Given the focus of this work on the automated, score-driven workflow we have selected cluster 1 as the docking outcome based on its superior scores, as provided by HADDOCK. Top model (model 1) from this cluster was refined using HADDOCK refinement interface and used in subsequent evaluation (Fig 2A).

Comparison of the NusA NTD/RNAP docking and experimental structural models

Coordinates for the experimental model of *E. coli* NusA-NTD were extracted from the cryo electron microscopy structure deposited as 6flq⁶¹, namely the chains C, D, and F for RpoB, RpoC, and NusA, respectively. Chain F was trimmed to the first 180 residues to match the size of NusA NTD used in docking simulation (Fig 2B). Superimposition (*super* command in PyMOL⁶²) revealed good overall congruity of these two models (Figure 2C), with NusA NTD binding in the same orientation to the same (beta flap) region of RNAP (consistent with other structural and biochemical data)^{46,63-65}. TM-score⁶⁶ of alignment, 0.4955, indicates near identity of the overall fold, and is very close to the TM-score of the aligned models without NusA NTD contribution (0.5450) (TM-scores of different experimental *E. coli* RNAP structure alignments vary between 0.5010 and 0.6205).

In order to ascertain the degree of agreement between the structural models (experimental and docking) and the experimental NusA NTD-RNAP X-links we have calculated the Euclidean CA-CA distances between X-linked residues in each structural model (Fig. 3). As can be seen in Table 2, 100% of distances in each case are compliant with the accepted range for DSS X-linker of <30 Å^{30,67}.

Altogether these findings strongly indicate that, although the docking model does not replicate the atomic details of the experimental (cryo EM) one, it sufficiently well reflects the binding poses and the overall architecture of the NusA-NTD/RNAP complex.

Discussion

We have executed and evaluated the results of an automated workflow for creating a structural model of a protein complex (*E. coli* NusA-NTD/RNAP), which was discovered and structurally interrogated *in vivo* by AP-XL-MS. The X-links-guided docking was performed without any additional input or bias (in fact the docking was completed prior to the time when the reference experimental structural model (6flq) became available). The entire workflow was constructed using freely available, user-friendly computational tools (pLink, HADDOCK, etc), and commercially available reagents. The entire computational pipeline is streamlined, featuring a minimal number of input/output hand-offs compared to other *in vivo* XL-MS approaches^{25,68}.

The efficiency of the workflow reported in this work approaches that of the top-performing DSSO-based methodology²⁵, while being more accessible to the laboratories lacking expertise necessary for its implementation. Utilizing broadly cell-permeable X-linker DSS, this workflow outperforms *in vivo* XL-MS approach based on the functionalized X-linker bis(succinimidyl)-3-azidomethyl-glutarate (BAMG)⁶⁸. Unlike DSS, BAMG is not readily available from commercial sources, exhibits restricted cell permeability (e.g. ineffective in treatment of *E. coli*) and toxicity, and, despite functionalization towards X-link enrichment and robust discovery⁶⁹, delivers low X-link discovery rates (84 inter-protein X-links from the entire *B. subtilis* proteome (4260 proteins))⁶⁸.

Targeted AP-XL-MS reported in this work in addition to the structural interrogation of the cell interactome during rapid growth of bacteria can be also applied to mammalian cells cultures, and the discovery of the conditional changes in PPC structure and composition in different physiological conditions (stationary phase, onset of virulence, heat shock, etc), or as response to administration of drugs and other bioactive compounds.

Materials and Methods

Buffer components and consumables

Proteomics-grade DSS and DMSO were acquired from Proteochem. Buffer components for cross-linking and protein purification were BioUltra grade (Millipore Sigma). LC-MS/MS was carried out with Thermo Scientific LC-MS grade reagents and solvents. Growth media components were from Thermo Fisher, protease inhibitor cocktail (ProBlock Gold Bacterial 2D) was from Gold Biotechnology, Ready-Lyse lysozyme solution was from Epicentre Biotechnologies, His-Mag Sepharose Ni – from GE. Low-binding pipet tips (Corning DeckWorks) and tubes (Protein LoBind, Eppendorf) were used throughout the experimental workflow.

In vivo X-linking and affinity purification of the RpoC-baited PPCs

E. coli strain RL721 (*rpoC::His6*) (generously contributed by Robert Landick, University of Wisconsin-Madison) was grown with agitation at 37°C in 0.5X phosphate-buffered Terrific Broth⁷⁰. DSS was dissolved in anhydrous DMSO (300 mM stock) and added to the bacterial culture at OD₆₀₀=0.5 at the final concentration of 2 mM. After 30 min incubation X-linking was quenched by the addition of Tris base to the final concentration of 50 mM. Cells were harvested by centrifugation (4°C, 10 min, 6 000 g), resuspended in lysis buffer (50 mM HEPES (pH 7.5), 500 mM NaCl, 1X ProBlock Gold Bacterial 2D) and lysed by the combined action of ultrasonication and Ready-Lyse lysozyme (30 KU/ml). Extract was cleared by centrifugation (4°C, 2X 30 min, 29 500 g) and combined with His-Mag Sepharose Ni, the mix was incubated at 4°C for 8 hrs with rotation. Unbound material was removed using magnetic separation, the beads were washed with denaturing buffer (8 M urea, 50 mM HEPES (pH 7.5), 500 mM NaCl) at 4°C for 2 hrs with rotation, twice with non-denaturing wash buffer (50 mM HEPES (pH 7.5), 500 mM NaCl) at 4°C for 2 hrs with rotation, the proteins were eluted in non-denaturing wash buffer, supplemented with 400 mM imidazole.

Processing of X-linked samples for and by LC-MS/MS was carried out by NYU Langone Health Proteomics Laboratory as described before⁷¹.

Discovery of the *in vivo* X-links

Discovery of the DSS X-links was carried out using pLink1³⁶. X-linked peptide search space was defined by combining protein sequences discovered in the enumerative analysis into a single *fasta* file. Search parameters were defined in the *pLink.ini* file by setting the enzyme name to trypsin, maximal number of missed cleavages to 3, maximal e-value to 0.001. Amino acid modifications were limited to 1 constant (Carbamidomethyl[C]), and 3 variable (Oxidation_M, Gln->pyro-Glu, and N-acetyl_Protein) ones. Example of the *pLink.ini* files is included in the Supplement.

XL-MS-guided protein-protein docking and analysis of the docking models

Distance restraints-guided protein-protein docking was carried out using the Expert interface of the HADDOCK server⁵⁹. Distance restraints were recorded in the *unambig.tbl* file. Starting structures and the *unambig.tbl* files are included in the Supplement. Homology modeling was carried out using the I-TASSER server⁵⁶. Starting structures refinement was carried out using FG-MD⁵⁷ and YASARA⁷² energy minimization servers, docking model refinement was performed using Refinement interface of the HADDOCK server. Coordinate files (*pdb*) manipulations (chain joining and renumbering) was carried out using YASARA Dynamics⁷³. Comparison between the top docking model and the reference experimental structure was carried out using TM-score server⁶⁶.

Table 1. Comparison of HADDOCK-generated metrics for NusA-NTD/RNAP models clusters 1 and 2

	Cluster 1	Cluster 2
HADDOCK score	127.0±9.0	152.0±4.1
Cluster size	117	81
RMSD from the lowest-energy structure	0.9±0.8	8.7±0.1
Van der Waals energy	-29.5±5.3	-27.2±9.8
Electrostatic energy	-551.8±96.7	-404.7±68.1
Desolvation energy	258.0±19.8	243.7±16.0
Restraints violation energy	87.9±6.99	165.5±27.32
Buried surface area	1901.2±134.6	1386.1±154.3
Z-score	-1.0	1.0

Table 2. Euclidean distances between CA-CA atoms in the docking and the experimental (cryo-EM) models of NusA-NTD/RNAP complex

	CA-CA distance (Å)	
	cryo-EM	docking
NusA3-RpoB900	13.9	23.6
NusA37-RpoB890	26.8	18.9
NusA37-RpoB909	22.2	6.2
NusA38-RpoB890	25.5	17.4
NusA38-RpoB909	20.7	6.8
NusA111-RpoB890	14.5	14.2
NusA111-RpoB909	17.8	12.5
NusA143-RpoC79	12.1	18.3

Figure 1. Scatter plot of the RMSD vs HADDOCK metrics for clusters 1 and 2. A: HADDOCK score; **B:** van der Waals energy; **C:** electrostatic energy, **D:** distance restraints violations (AIRs).

Figure 2. Structures of NusA-NTD/RNAP complex. A: docking model (RNAP=blue cartoon, NusA-NTD=orange cartoon); **B:** cryo EM model (RNAP=teal cartoon, NusA-NTD=hot pink cartoon); **C:** superimposed docking (A) and cryo EM(B) models.

Figure 3. *In vivo* X-linked residues mapped on the docking and cryo EM models of NusA-NTD/RNAP. Structures are pictured as semi-transparent cartoons, CA atoms in X-linked residues – as solid spheres; color scheme is the same as in Fig. 2.

References

- Tran, B. Q., Goodlett, D. R. & Goo, Y. A. Advances in protein complex analysis by chemical cross-linking coupled with mass spectrometry (CXMS) and bioinformatics. *Biochim Biophys Acta* **1864**, 123-129, doi:10.1016/j.bbapap.2015.05.015 (2016).
- Schneider, M., Belsom, A. & Rappsilber, J. Protein Tertiary Structure by Crosslinking/Mass Spectrometry. *Trends Biochem Sci* **43**, 157-169, doi:10.1016/j.tibs.2017.12.006 (2018).
- Leitner, A., Faini, M., Stengel, F. & Aebersold, R. Crosslinking and Mass Spectrometry: An Integrated Technology to Understand the Structure and Function of Molecular Machines. *Trends Biochem Sci* **41**, 20-32, doi:10.1016/j.tibs.2015.10.008 (2016).
- Liu, F. & Heck, A. J. Interrogating the architecture of protein assemblies and protein interaction networks by cross-linking mass spectrometry. *Curr Opin Struct Biol* **35**, 100-108, doi:10.1016/j.sbi.2015.10.006 (2015).
- Bullock, J. M. A., Sen, N., Thalassinou, K. & Topf, M. Modeling Protein Complexes Using Restraints from Crosslinking Mass Spectrometry. *Structure*, doi:10.1016/j.str.2018.04.016 (2018).
- Sinz, A. The advancement of chemical cross-linking and mass spectrometry for structural proteomics: from single proteins to protein interaction networks. *Expert Rev Proteomics* **11**, 733-743, doi:10.1586/14789450.2014.960852 (2014).
- Aebersold, R. & Mann, M. Mass-spectrometric exploration of proteome structure and function. *Nature* **537**, 347-355, doi:10.1038/nature19949 (2016).
- Faini, M., Stengel, F. & Aebersold, R. The Evolving Contribution of Mass Spectrometry to Integrative Structural Biology. *J Am Soc Mass Spectrom* **27**, 966-974, doi:10.1007/s13361-016-1382-4 (2016).
- Chait, B. T., Cadene, M., Olinares, P. D., Rout, M. P. & Shi, Y. Revealing Higher Order Protein Structure Using Mass Spectrometry. *J Am Soc Mass Spectrom* **27**, 952-965, doi:10.1007/s13361-016-1385-1 (2016).
- Leitner, A., Walzthoeni, T. & Aebersold, R. Lysine-specific chemical cross-linking of protein complexes and identification of cross-linking sites using LC-MS/MS and the xQuest/xProphet software pipeline. *Nat Protoc* **9**, 120-137, doi:10.1038/nprot.2013.168 (2014).

- 11 Kalkhof, S. & Sinz, A. Chances and pitfalls of chemical cross-linking with amine-reactive N-hydroxysuccinimide esters. *Anal Bioanal Chem* **392**, 305-312 (2008).
- 12 Madler, S., Bich, C., Touboul, D. & Zenobi, R. Chemical cross-linking with NHS esters: a systematic study on amino acid reactivities. *J Mass Spectrom* **44**, 694-706, doi:10.1002/jms.1544 (2009).
- 13 Madler, S., Gschwind, S. & Zenobi, R. Role of arginine in chemical cross-linking with N-hydroxysuccinimide esters. *Anal Biochem* **398**, 123-125, doi:10.1016/j.ab.2009.11.020 (2010).
- 14 Wong, S. S., Jameson, D. M. & Wong, S. S. *Chemistry of protein and nucleic acid cross-linking and conjugation*. 2nd edn, (Taylor & Francis/CRC Press, 2012).
- 15 Schmidt, C. & Urlaub, H. Combining cryo-electron microscopy (cryo-EM) and cross-linking mass spectrometry (CX-MS) for structural elucidation of large protein assemblies. *Curr Opin Struct Biol* **46**, 157-168, doi:10.1016/j.sbi.2017.10.005 (2017).
- 16 Demo, G. et al. Structure of RNA polymerase bound to ribosomal 30S subunit. *Elife* **6**, doi:10.7554/eLife.28560 (2017).
- 17 Bodnar, N. O. et al. Structure of the Cdc48 ATPase with its ubiquitin-binding cofactor Ufd1-Npl4. *Nature Structural & Molecular Biology*, doi:10.1038/s41594-018-0085-x (2018).
- 18 Srinivasa, S., Ding, X. & Kast, J. Formaldehyde cross-linking and structural proteomics: Bridging the gap. *Methods* **89**, 91-98, doi:10.1016/j.ymeth.2015.05.006 (2015).
- 19 Wierer, M. & Mann, M. Proteomics to study DNA-bound and chromatin-associated gene regulatory complexes. *Hum Mol Genet* **25**, R106-R114, doi:10.1093/hmg/ddw208 (2016).
- 20 Gustafsson, O. J., Arentz, G. & Hoffmann, P. Proteomic developments in the analysis of formalin-fixed tissue. *Biochim Biophys Acta* **1854**, 559-580, doi:10.1016/j.bbapap.2014.10.003 (2015).
- 21 Klockenbusch, C., O'Hara, J. E. & Kast, J. Advancing formaldehyde cross-linking towards quantitative proteomic applications. *Anal Bioanal Chem* **404**, 1057-1067, doi:10.1007/s00216-012-6065-9 (2012).
- 22 Varjosalo, M. et al. Interlaboratory reproducibility of large-scale human protein-complex analysis by standardized AP-MS. *Nat Methods* **10**, 307-314, doi:10.1038/nmeth.2400 (2013).
- 23 Keilhauer, E. C., Hein, M. Y. & Mann, M. Accurate protein complex retrieval by affinity enrichment mass spectrometry (AE-MS) rather than affinity purification mass spectrometry (AP-MS). *Mol Cell Proteomics* **14**, 120-135, doi:10.1074/mcp.M114.041012 (2015).
- 24 Rajala, N. et al. Whole cell formaldehyde cross-linking simplifies purification of mitochondrial nucleoids and associated proteins involved in mitochondrial gene expression. *PLoS One* **10**, e0116726, doi:10.1371/journal.pone.0116726 (2015).
- 25 Wang, X. et al. Molecular Details Underlying Dynamic Structures and Regulation of the Human 26S Proteasome. *Mol Cell Proteomics* **16**, 840-854, doi:10.1074/mcp.M116.065326 (2017).
- 26 Chen, Z. A. et al. Architecture of the RNA polymerase II-TFIIF complex revealed by cross-linking and mass spectrometry. *EMBO J* **29**, 717-726, doi:10.1038/emboj.2009.401 (2010).
- 27 Kahraman, A. et al. Cross-link guided molecular modeling with ROSETTA. *PLoS One* **8**, e73411, doi:10.1371/journal.pone.0073411 (2013).
- 28 Debelyy, M. O., Waridel, P., Quadroni, M., Schneiter, R. & Conzelmann, A. Chemical crosslinking and mass spectrometry to elucidate the topology of integral membrane proteins. *PLoS One* **12**, e0186840, doi:10.1371/journal.pone.0186840 (2017).
- 29 Makowski, M. M., Willems, E., Jansen, P. W. & Vermeulen, M. Cross-linking immunoprecipitation-MS (xIP-MS): Topological Analysis of Chromatin-associated Protein Complexes Using Single Affinity Purification. *Mol Cell Proteomics* **15**, 854-865, doi:10.1074/mcp.M115.053082 (2016).
- 30 Merkley, E. D. et al. Distance restraints from crosslinking mass spectrometry: mining a molecular dynamics simulation database to evaluate lysine-lysine distances. *Protein Sci* **23**, 747-759, doi:10.1002/pro.2458 (2014).
- 31 Bartels, T., Choi, J. G. & Selkoe, D. J. alpha-Synuclein occurs physiologically as a helically folded tetramer that resists aggregation. *Nature* **477**, 107-110, doi:10.1038/nature10324 (2011).
- 32 Huber, L. J. & Chao, M. V. A potential interaction of p75 and trkA NGF receptors revealed by affinity crosslinking and immunoprecipitation. *J Neurosci Res* **40**, 557-563, doi:10.1002/jnr.490400415 (1995).
- 33 Zeng, P. Y., Vakoc, C. R., Chen, Z. C., Blobel, G. A. & Berger, S. L. In vivo dual cross-linking for identification of indirect DNA-associated proteins by chromatin immunoprecipitation. *Biotechniques* **41**, 694, 696, 698, doi:10.2144/000112297 (2006).

- 34 Lunardi-Iskandar, Y. *et al.* Tumorigenesis and metastasis of neoplastic Kaposi's sarcoma cell line in immunodeficient mice blocked by a human pregnancy hormone. *Nature* **375**, 64-68, doi:10.1038/375064a0 (1995).
- 35 Trupp, M. *et al.* Functional receptor for GDNF encoded by the c-ret proto-oncogene. *Nature* **381**, 785-789, doi:10.1038/381785a0 (1996).
- 36 Yang, B. *et al.* Identification of cross-linked peptides from complex samples. *Nat Methods* **9**, 904-906, doi:10.1038/nmeth.2099 (2012).
- 37 Mellacheruvu, D. *et al.* The CRAPome: a contaminant repository for affinity purification-mass spectrometry data. *Nat Methods* **10**, 730-736, doi:10.1038/nmeth.2557 (2013).
- 38 Schmidt, A. *et al.* The quantitative and condition-dependent Escherichia coli proteome. *Nat Biotechnol* **34**, 104-110, doi:10.1038/nbt.3418 (2016).
- 39 Belogurov, G. A. & Artsimovitch, I. Regulation of Transcript Elongation. *Annu Rev Microbiol* **69**, 49-69, doi:10.1146/annurev-micro-091014-104047 (2015).
- 40 Yamamoto, K. The hierarchic network of metal-response transcription factors in Escherichia coli. *Biosci Biotechnol Biochem* **78**, 737-747, doi:10.1080/09168451.2014.915731 (2014).
- 41 Yang, X. & Lewis, P. J. The interaction between bacterial transcription factors and RNA polymerase during the transition from initiation to elongation. *Transcription* **1**, 66-69, doi:10.4161/trns.1.2.12791 (2010).
- 42 Mooney, R. A. *et al.* Regulator trafficking on bacterial transcription units in vivo. *Mol Cell* **33**, 97-108, doi:10.1016/j.molcel.2008.12.021 (2009).
- 43 Strauss, M. *et al.* Transcription is regulated by NusA:NusG interaction. *Nucleic Acids Res* **44**, 5971-5982, doi:10.1093/nar/gkw423 (2016).
- 44 Mondal, S., Yakhnin, A. V., Sebastian, A., Albert, I. & Babitzke, P. NusA-dependent transcription termination prevents misregulation of global gene expression. *Nat Microbiol* **1**, 15007, doi:10.1038/nmicrobiol.2015.7 (2016).
- 45 Wells, C. D., Deighan, P., Brigham, M. & Hochschild, A. Nascent RNA length dictates opposing effects of NusA on antitermination. *Nucleic Acids Res* **44**, 5378-5389, doi:10.1093/nar/gkw198 (2016).
- 46 Ma, C. *et al.* RNA polymerase-induced remodelling of NusA produces a pause enhancement complex. *Nucleic Acids Res* **43**, 2829-2840, doi:10.1093/nar/gkv108 (2015).
- 47 Zhou, J., Ha, K. S., La Porta, A., Landick, R. & Block, S. M. Applied force provides insight into transcriptional pausing and its modulation by transcription factor NusA. *Mol Cell* **44**, 635-646, doi:10.1016/j.molcel.2011.09.018 (2011).
- 48 Schmidt, M. C. & Chamberlin, M. J. Binding of rho factor to Escherichia coli RNA polymerase mediated by nusA protein. *J Biol Chem* **259**, 15000-15002 (1984).
- 49 Schmidt, M. C. & Chamberlin, M. J. Amplification and isolation of Escherichia coli nusA protein and studies of its effects on in vitro RNA chain elongation. *Biochemistry* **23**, 197-203 (1984).
- 50 Schmidt, M. C. & Chamberlin, M. J. nusA protein of Escherichia coli is an efficient transcription termination factor for certain terminator sites. *J Mol Biol* **195**, 809-818 (1987).
- 51 Cardinale, C. J. *et al.* Termination factor Rho and its cofactors NusA and NusG silence foreign DNA in E. coli. *Science* **320**, 935-938, doi:10.1126/science.1152763 (2008).
- 52 Burns, C. M., Richardson, L. V. & Richardson, J. P. Combinatorial effects of NusA and NusG on transcription elongation and Rho-dependent termination in Escherichia coli. *J Mol Biol* **278**, 307-316, doi:10.1006/jmbi.1998.1691 (1998).
- 53 Shankar, S., Hatoum, A. & Roberts, J. W. A transcription antiterminator constructs a NusA-dependent shield to the emerging transcript. *Mol Cell* **27**, 914-927, doi:10.1016/j.molcel.2007.07.025 (2007).
- 54 Cohen, S. E. *et al.* Roles for the transcription elongation factor NusA in both DNA repair and damage tolerance pathways in Escherichia coli. *Proc Natl Acad Sci U S A* **107**, 15517-15522, doi:10.1073/pnas.1005203107 (2010).
- 55 Karaca, E., Melquiond, A. S., de Vries, S. J., Kastitis, P. L. & Bonvin, A. M. Building macromolecular assemblies by information-driven docking: introducing the HADDOCK multibody docking server. *Mol Cell Proteomics* **9**, 1784-1794, doi:10.1074/mcp.M000051-MCP201 (2010).
- 56 Yang, J. & Zhang, Y. Protein Structure and Function Prediction Using I-TASSER. *Curr Protoc Bioinformatics* **52**, 5 8 1-15, doi:10.1002/0471250953.bi0508s52 (2015).

- 57 Xu, D., Zhang, J., Roy, A. & Zhang, Y. Automated protein structure modeling in CASP9 by I-TASSER pipeline combined with QUARK-based ab initio folding and FG-MD-based structure refinement. *Proteins* **79 Suppl 10**, 147-160, doi:10.1002/prot.23111 (2011).
- 58 Bae, B. *et al.* Phage T7 Gp2 inhibition of Escherichia coli RNA polymerase involves misappropriation of sigma70 domain 1.1. *Proc Natl Acad Sci U S A* **110**, 19772-19777, doi:10.1073/pnas.1314576110 (2013).
- 59 van Zundert, G. C. & Bonvin, A. M. Modeling protein-protein complexes using the HADDOCK webserver "modeling protein complexes with HADDOCK". *Methods Mol Biol* **1137**, 163-179, doi:10.1007/978-1-4939-0366-5_12 (2014).
- 60 Bonvin, A., Karaca, E., Kastiris, P. L. & Rodrigues, J. Defining distance restraints in HADDOCK. *Nat Protoc* **13**, 1503, doi:10.1038/s41596-018-0017-6 (2018).
- 61 Guo, X. *et al.* Structural Basis for NusA Stabilized Transcriptional Pausing. *Mol Cell* **69**, 816-827 e814, doi:10.1016/j.molcel.2018.02.008 (2018).
- 62 Yuan, S., Chan, H. C. S., Filipek, S. & Vogel, H. PyMOL and Inkscape Bridge the Data and the Data Visualization. *Structure* **24**, 2041-2042, doi:10.1016/j.str.2016.11.012 (2016).
- 63 Yang, X. *et al.* The structure of bacterial RNA polymerase in complex with the essential transcription elongation factor NusA. *EMBO Rep* **10**, 997-1002, doi:10.1038/embor.2009.155 (2009).
- 64 Drogemuller, J. *et al.* Determination of RNA polymerase binding surfaces of transcription factors by NMR spectroscopy. *Sci Rep* **5**, 16428, doi:10.1038/srep16428 (2015).
- 65 Bylund, G. O., Nord, S., Lovgren, J. M. & Wikstrom, P. M. Alterations in the beta flap and beta' dock domains of the RNA polymerase abolish NusA-mediated feedback regulation of the metY-nusA-infB operon. *J Bacteriol* **193**, 4113-4122, doi:10.1128/JB.00196-11 (2011).
- 66 Zhang, Y. & Skolnick, J. Scoring function for automated assessment of protein structure template quality. *Proteins* **57**, 702-710, doi:10.1002/prot.20264 (2004).
- 67 Kahraman, A., Malmstrom, L. & Aebersold, R. Xwalk: computing and visualizing distances in cross-linking experiments. *Bioinformatics* **27**, 2163-2164, doi:10.1093/bioinformatics/btr348 (2011).
- 68 de Jong, L. *et al.* In-Culture Cross-Linking of Bacterial Cells Reveals Large-Scale Dynamic Protein-Protein Interactions at the Peptide Level. *J Proteome Res* **16**, 2457-2471, doi:10.1021/acs.jproteome.7b00068 (2017).
- 69 Buncherd, H. *et al.* Selective enrichment and identification of cross-linked peptides to study 3-D structures of protein complexes by mass spectrometry. *J Proteomics* **75**, 2205-2215, doi:10.1016/j.jprot.2012.01.025 (2012).
- 70 Studier, F. W. Stable expression clones and auto-induction for protein production in E. coli. *Methods Mol Biol* **1091**, 17-32, doi:10.1007/978-1-62703-691-7_2 (2014).
- 71 Epshtein, V. *et al.* UvrD facilitates DNA repair by pulling RNA polymerase backwards. *Nature* **505**, 372-377, doi:10.1038/nature12928 (2014).
- 72 Krieger, E. *et al.* Improving physical realism, stereochemistry, and side-chain accuracy in homology modeling: Four approaches that performed well in CASP8. *Proteins* **77 Suppl 9**, 114-122, doi:10.1002/prot.22570 (2009).
- 73 Krieger, E. & Vriend, G. YASARA View - molecular graphics for all devices - from smartphones to workstations. *Bioinformatics* **30**, 2981-2982, doi:10.1093/bioinformatics/btu426 (2014).

Acknowledgements

This work was supported by the NIH grant R01 GM107329 and by the Howard Hughes Medical Institute.

Competing interests

None declared.

Contributions

EN and VS conceived the study. VS carried out the experiments and data analysis under EN supervision. EN and VS wrote the manuscript.

Materials and Correspondence

All correspondence and requests for materials should be submitted to Evgeny Nudler (Evgeny.Nudler@nyumc.org). Technical questions may also be directed to Vladimir Svetlov (Vladimir.Svetlov@nyumc.org).

Figure 1A

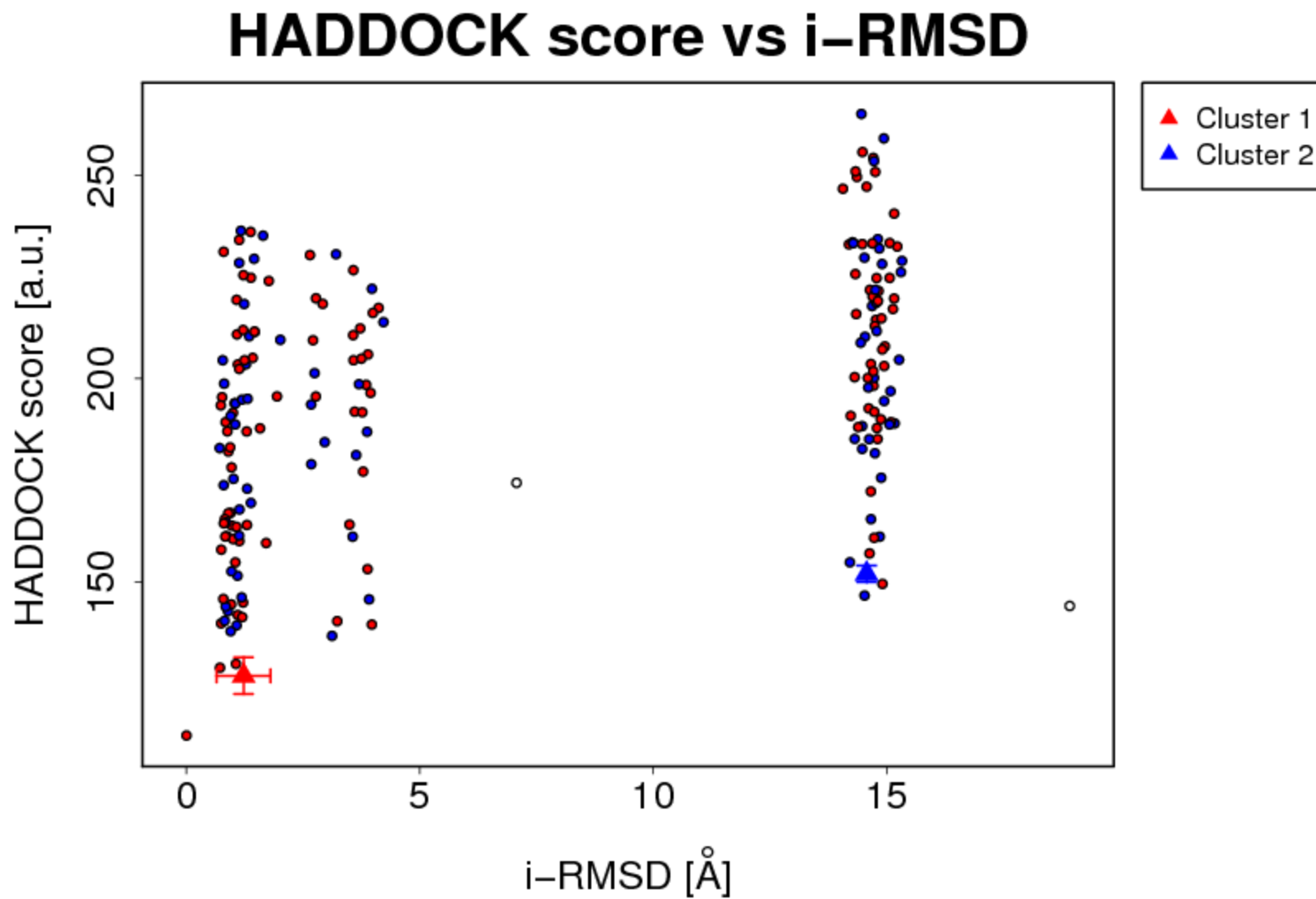


Figure 1B

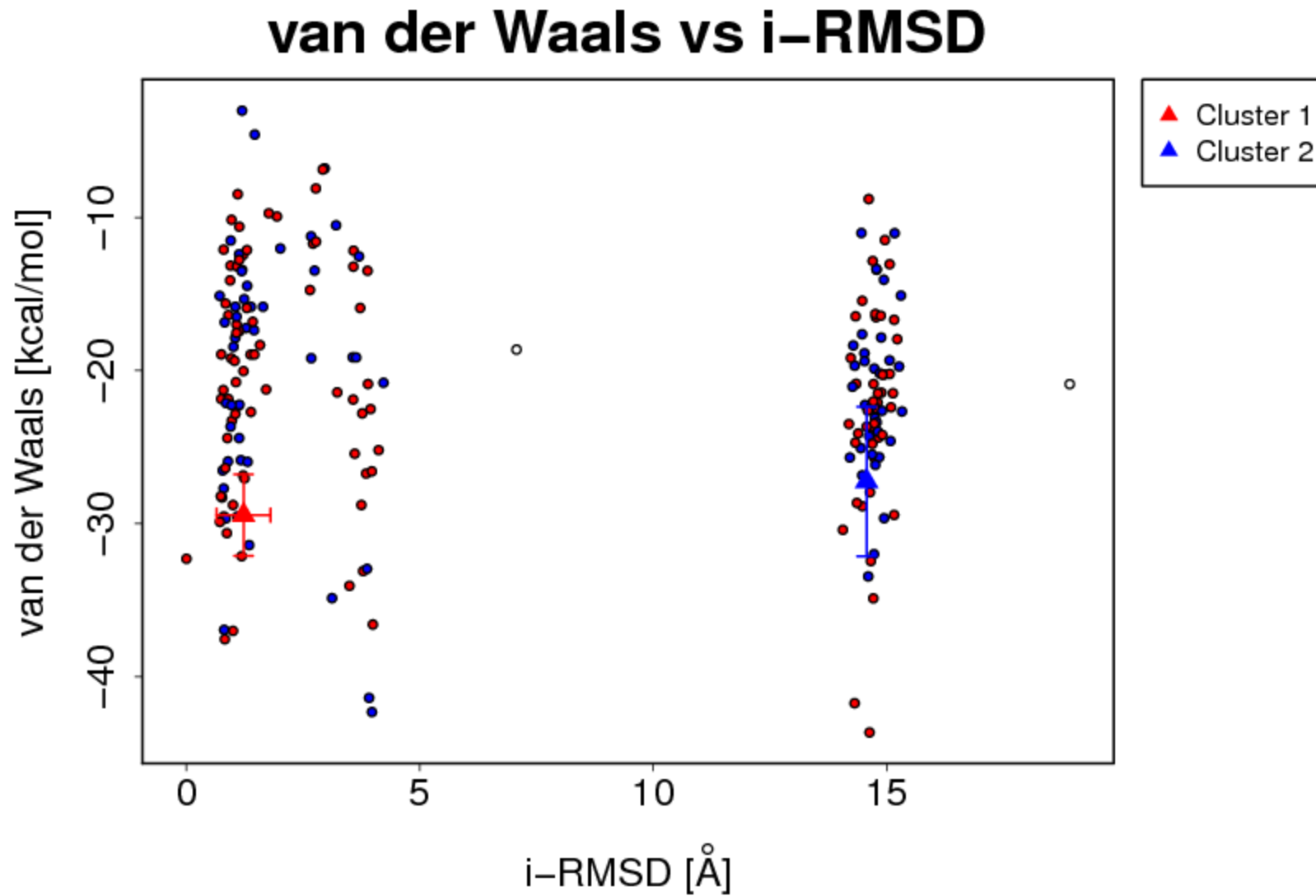


Figure 1C

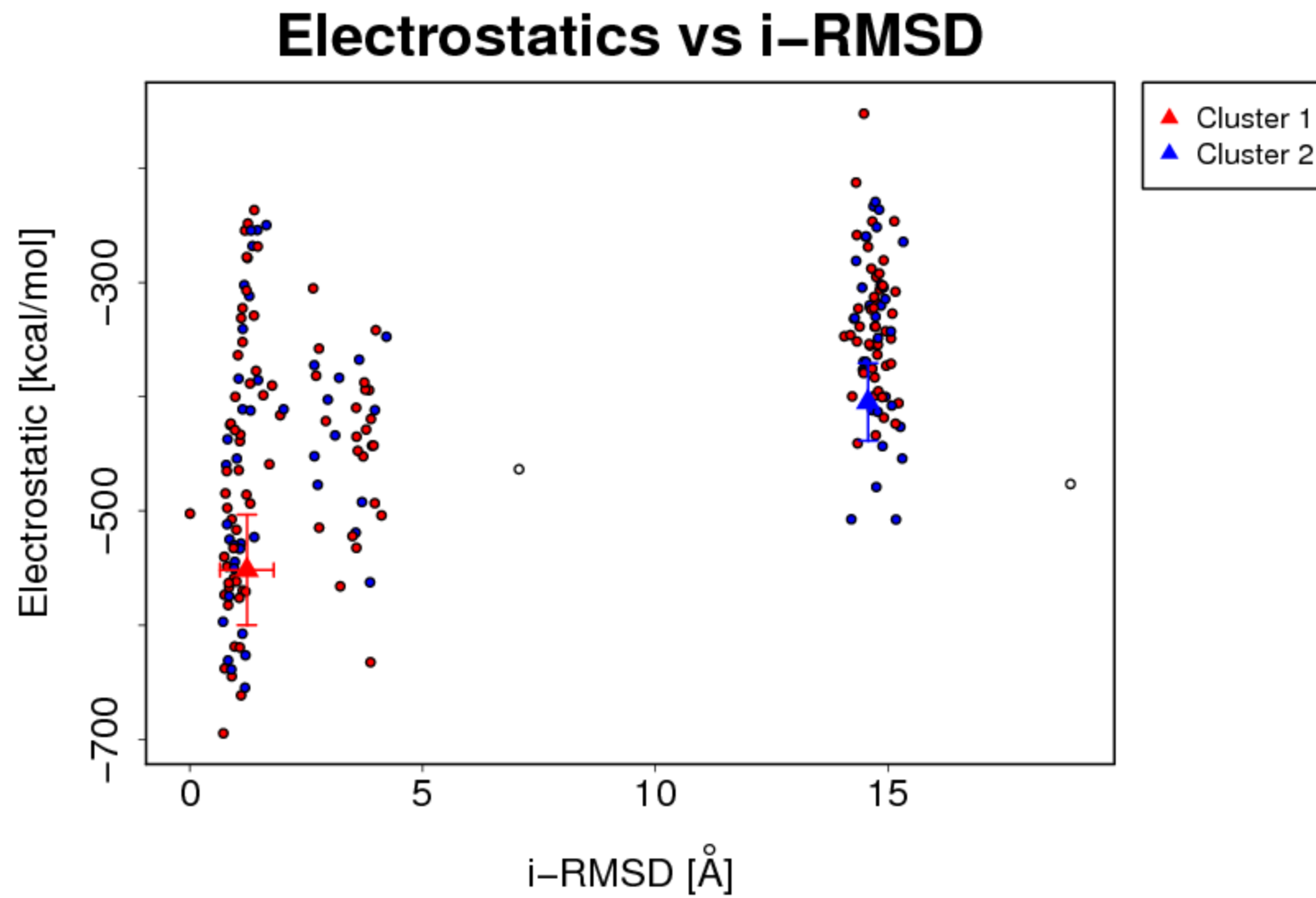


Figure 1D

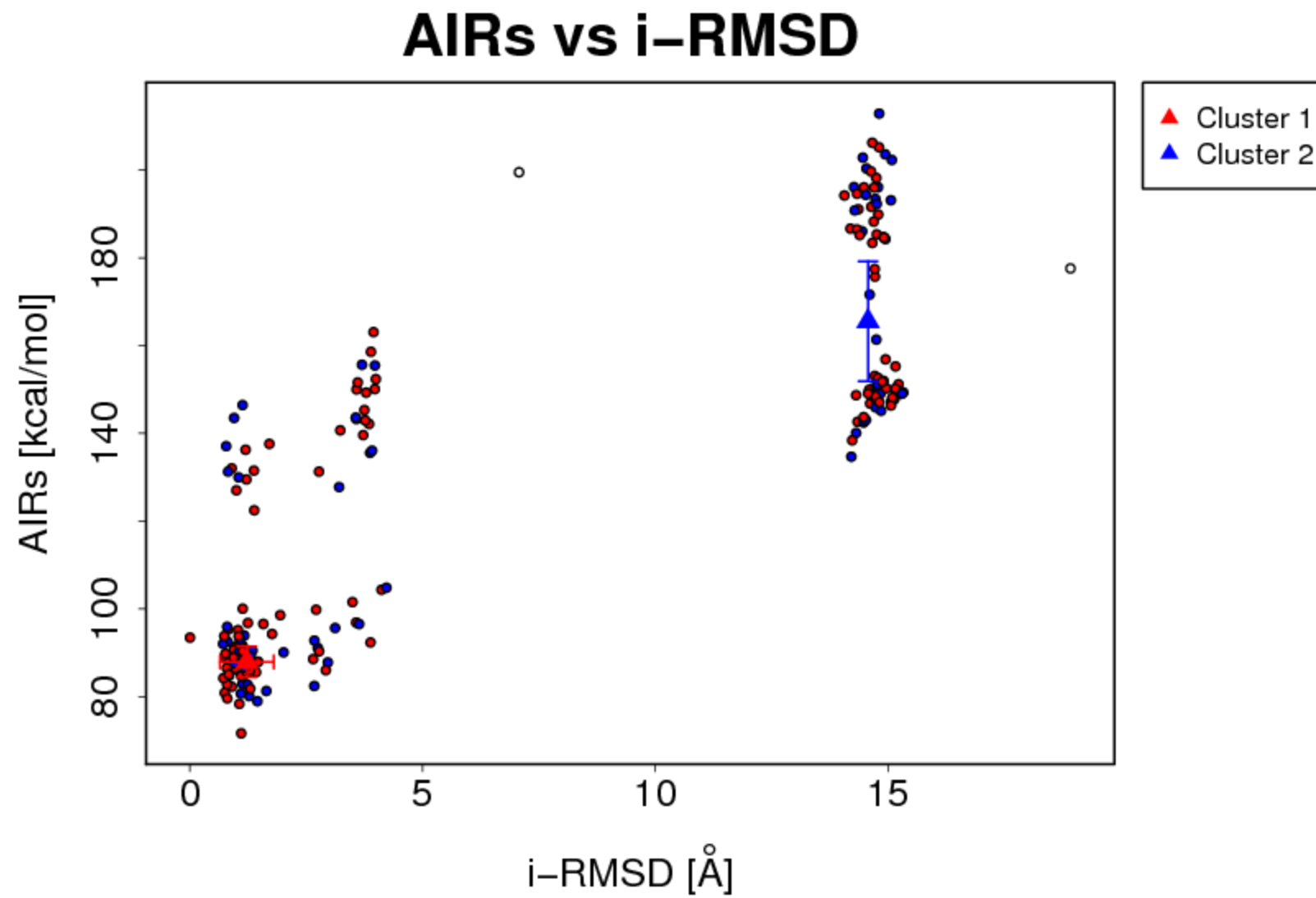


Figure 2A

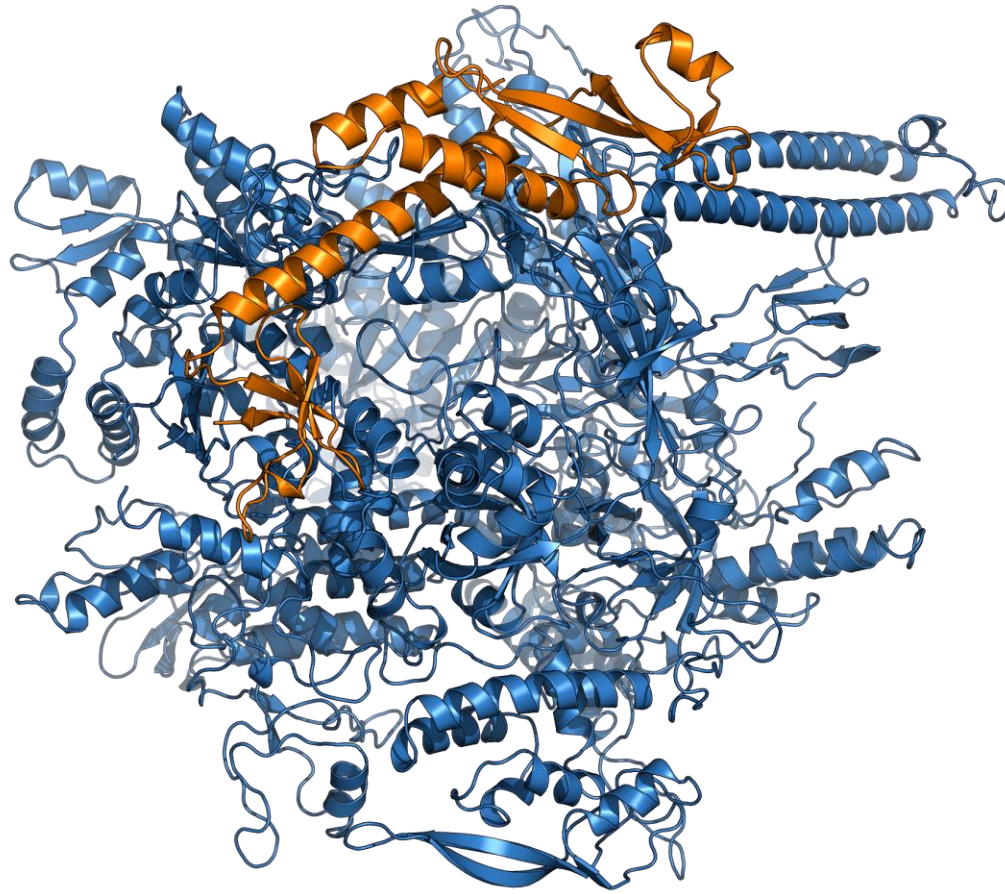


Figure 2B

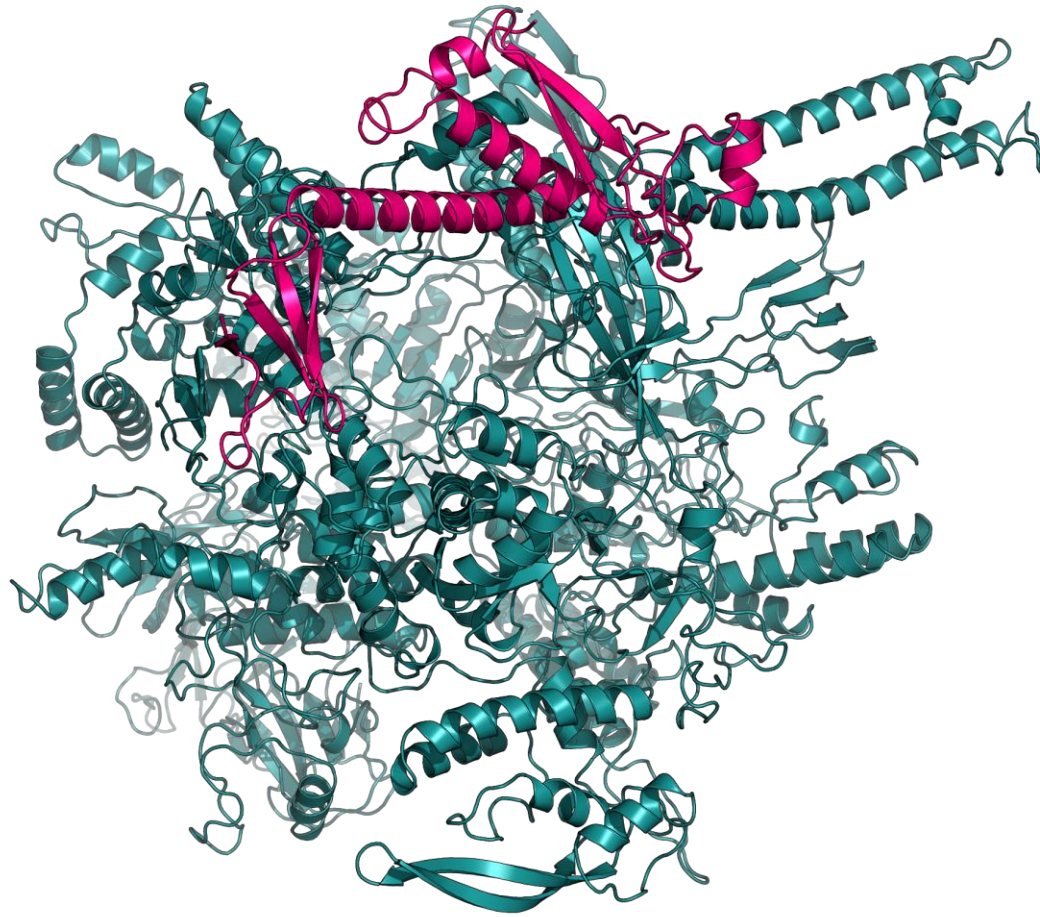


Figure 2C

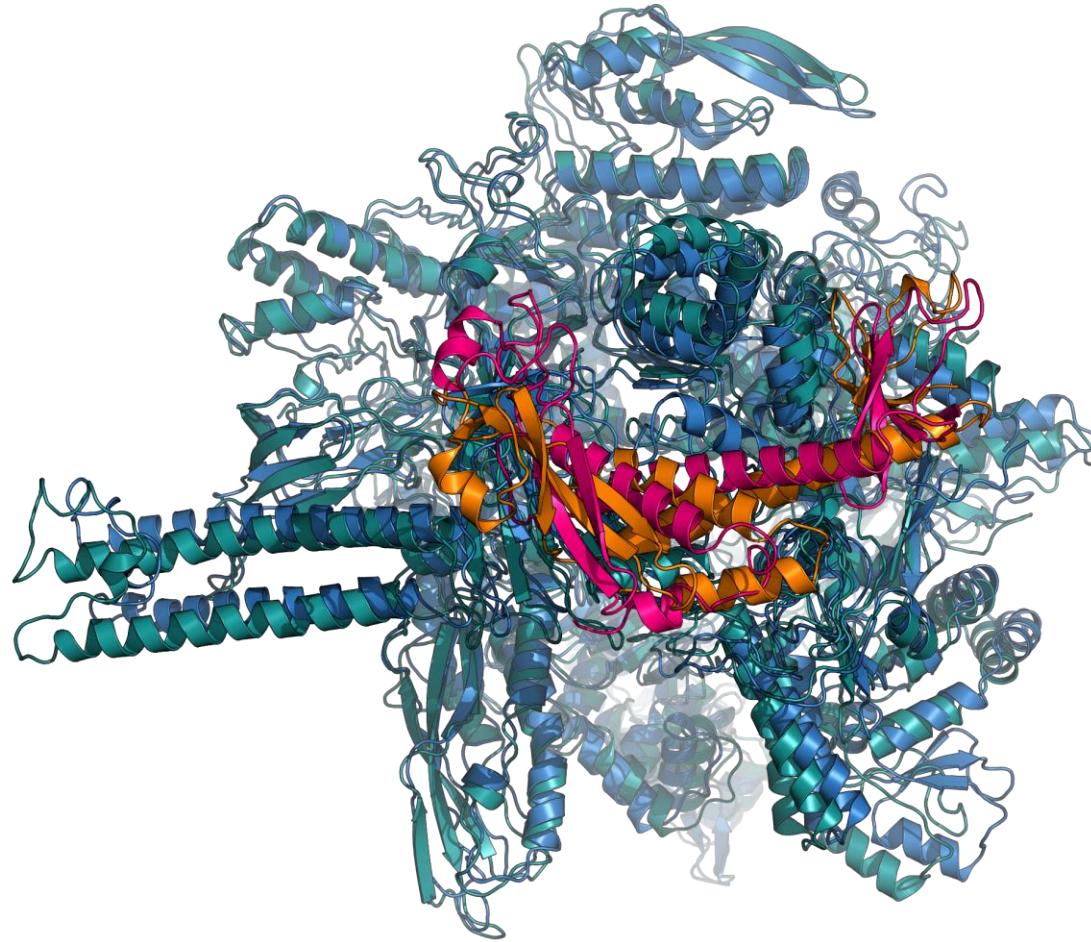


Figure 3

