High-fidelity mass analysis unveils heterogeneity in intact ribosomal particles

Michiel van de Waterbeemd[1,2*], Kyle L Fort[1,2*], Dmitriy Boll[3], Maria Reinhardt-Szyba[3], Andrew Routh[4,5], Alexander Makarov[1,3], Albert J R Heck[1,2#]

* These authors contributed equally
# Corresponding author

[1] Biomolecular Mass Spectrometry and Proteomics, Bijvoet Center for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, Utrecht University, 3584 CH, Utrecht, The Netherlands.
[2] Netherlands Proteomics Center, Padualaan 8, Utrecht, 3584 CH, The Netherlands.
[3] Thermo Fisher Scientific (Bremen), 28199 Bremen, Germany.
[4] Department of Biochemistry and Molecular Biology, University of Texas Medical Branch, Galveston, Texas, USA.
[5] Sealy Centre for Structural Biology and Molecular Biophysics, University of Texas Medical Branch, Galveston, Texas, USA.

Investigation of the structure, assembly and function of protein-nucleic acid macromolecular machines requires multidimensional approaches in both molecular and structural biology. Here we describe new modifications to an Orbitrap mass spectrometer, enabling the high resolution native MS analysis of 0.8-2.3 MDa prokaryotic 30S, 50S and 70S ribosome particles and the 9 MDa Flock House virus. The instrument's improved mass range and sensitivity readily exposes the unexpected binding of the ribosome associated protein *sra*.

The use of hybrid approaches in structural biology, for instance cryo-electron microscopy and crosslinking mass spectrometry, has yielded a tremendous amount of new insight in the workings of some of nature's most intricate molecular machines[1–6]. Native Mass Spectrometry (native MS)[7] can play an eminent role in the characterization of these molecular machines by providing insight into their composition, modularity and ligand binding[8–10], revealing structural details that might be elusive using other methods. However, owing to their high complexity and molecular weight, analysis of these protein complexes requires a dedicated mass spectrometer that combines high mass resolution, high mass range and high sensitivity. The recently introduced Extended Mass Range Orbitrap™ (EMR) instrument[11] was designed to address this need and has now been successfully used for systems in the 20-250 kilo-Dalton (kDa) range[12,13]. Although this instrument could also successfully analyze whole viruses up to 4.5 mega-Dalton (MDa), its low sensitivity at high mass to charge ratios (*m/z)* greatly reduced its applicability in analyzing such particles[14,15].

Protein–nucleic acid complexes attain a relatively low number of charges during electrospray ionization as their oligonucleotide components contribute to the mass but maintain little to no net positive charge (Figure 1a). This property of protein-nucleic acid complexes, together with their high molecular weight, make them particularly challenging to analyze with current native mass spectrometry instrumentation. Therefore, we set out to modify an Orbitrap instrument with the aim to particularly improve the sensitivity at high *m/z*, thereby extending the upper mass limit and enabling high-resolution accurate mass measurement of large protein–nucleic acid complexes. Distinguishingly, the QE-UHMR (Q-Exactive Ultra-High Mass Range), as it is named, is based on the Q-Exactive platform, but uniquely features a high mass quadrupole and several modifications enhancing the efficient detection

of ions of very high *m/z*. To demonstrate its capabilities, we analyzed multiple high molecular weight protein–nucleic acid assemblies: 30S, 50S and 70S *E. coli* ribosomal particles and authentic Flock House virus particles containing genomic RNA.

Critical to successful native mass spectrometric analysis of large protein-nucleic acid complexes is the efficient transmission of their ions, which exhibit very high *m/z* values, through the mass spectrometer to the detector. While the Orbitrap mass analyzer has no theoretical upper *m/z* limit, transmission through radio frequency (RF) confining multipoles greatly hampers the detection of high *m/z* ions. Our modifications (methods section), are aimed at increasing the transmission and resolution of ions with an *m/z* ratio above 20,000. Briefly, focusing of ions is enhanced by lowering the RF frequency of the instrument's ion guides. In-source trapping of the ions provides improved and controllable desolvation while focusing the ion cloud in the front end of the mass spectrometer. Finally, transport of the high *m/z* ions from the C-trap into the Orbitrap mass analyzer is optimized to efficiently capture them (Supplementary Fig. 1). The collective result of these modifications is a substantial increase in ion transmission and an extension of the upper experimental *m/z* limit to more than double that compared to the EMR, as can be shown through analysis of Hepatitis B virus like particles (Figure 1B, Supplementary Fig. 2a). Additionally, the high mass quadrupole of the QE-UHMR allows isolation and fragmentation of protein complexes at very high *m/z* values, enabling tandem MS analysis (Supplementary Fig. 2b-c)[16]. Here, we demonstrate the enhanced capabilities of this instrument through the analysis of several challenging protein–nucleic acid assemblies.

Prokaryotic ribosomes are large, protein-ribonucleic acid molecular machines responsible for the translation of proteins from messenger RNA transcripts. They consist of two parts: the smaller 30S ribosome and the larger 50S ribosome that assemble together in the presence of Mg^{2+} ions into the 70S ribosome[16]. The 30S ribosomes are mainly responsible for translation initiation and consist of 21 proteins (RS1 to RS21) and a fragment of ribosomal RNA (16S rRNA). The 50S ribosome, which contains the catalytic peptidyl transferase site, consists of 36 proteins (RL1 to RL36) and two rRNA fragments (5S and 23S rRNA). The complete 70S ribosome reaches a molecular weight of approximately 2.30 mega-Dalton[17]. Structural investigation of *E. coli* ribosomes is challenging, not only due to the size and molecular complexity of these molecular machines, but also due to their inherent heterogeneity. Some ribosomal proteins may be present sub-stoichiometrically while other ribosome interacting proteins may be recruited at the different stages of translation. Additionally, small modifications have been mapped to both the ribosomal proteins and rRNA fragments[18,19]. A further challenge for native MS analysis is the requirement of Mg^{2+} ions to stabilize in particular the 70S particles, which causes significant peak broadening and signal suppression in the resulting mass spectra. Notwithstanding the fact that the Robinson group reported impressive MS data on ribosome particles already a decade ago[20,21], mass spectra displaying the heterogeneity in the 30S and 50S particles are absent and previously no mass could be assigned from the unresolved spectra of the 70S particles.

Here, we demonstrate a profound advancement in the native MS analysis of 70S, 30S and 50S ribosome particles through the use of the newly developed QE-UHMR mass spectrometer. The enhanced transmission for high *m/z* ions allows detection of high molecular weight complexes. Moreover, the gain in sensitivity allows the use of charge reduction through the addition of triethylammonium acetate to the spray solution in order to further enhance the mass resolving power[15,22]. Figure 2a (top panel) shows the mass spectrum of intact *E. coli* 70S ribosomes in the presence of 10 mM Mg^{2+}, with ions centered around 36,000 *m/z*. The well-resolved series of charge states allows determination of the accurate mass of the fully assembled 70S ribosome of 2316 ± 1 kDa, which matches well with the expected mass (2302 kDa). The mass deviation of 0.6% with the theoretical mass is partly attributable to incomplete desolvation but is mainly caused by the non-specific adduction of the Mg^{2+} ions.

Lowering the Mg^{2+} concentration leads to disassembly of the 70S particles into the 30S and 50S particles (Figure 2a, bottom panel). Using the QE-UHMR, these native MS spectra revealed a number of previously hidden details. The two most abundant 30S ribosome assemblies (788.59 ± 0.08 kDa and 850.1 ± 0.10 kDa) exhibit a mass difference of approximately 61.50 kDa (figure 2B). The mass of 850.09 kDa is in good correspondence with the expected mass for complete 30S particles (847.54 kDa) and we hypothesize that the mass of 788.59 kDa originates

from the sub-stoichiometric presence of the RS1 protein (expected mass loss 61.2 kDa). The RS1 protein is loosely associated to the 30S particles and is highly dynamic. Therefore, the protein is often intentionally removed before crystallization and is absent in all high resolution cryo-electron microscopy reconstructions[23]. RS1 mediates the interaction of the mRNA with the 16S rRNA and is essential for *E. coli* translation *in vivo*[24,25]. We additionally resolved two lower abundant 30S particles (793.75 ± 0.08 and 855.26 ± 0.05 kDa) that also display the partial absence of the RS1 protein. However, the 5.16 kDa mass difference between the two sets of assemblies seems to indicate the presence of a small ribosome interacting protein. Top-down tandem MS experiments of the 30S ribosome particles (Figure 2d and Supplementary Fig. 3) provided a more accurate mass of 5095 Da which we assigned to the 'stationary-phase-induced ribosomal associated protein' (*sra,* expected mass 5095.8 Da)[18]. Bottom-up LC-MS/MS analysis of all proteins present in the 70S ribosome preparation confirmed the presence of the *sra* protein and revealed it to be the most abundant non-ribosomal protein in the preparation (Figure 2e). The *sra* protein is a ribosome interactor that is bound to around 10% of the *E. coli* ribosomes during log phase growth, but this increases to 40% when the bacteria enter stationary phase[26]. Our measurements indicate that approximately 22% of the 30S particles in the preparation have *sra* protein bound.

The majority of the significantly larger 50S ribosome particles (Figure 2c) exhibited a mass of 1389.7 ± 0.15 kDa, which is in good agreement with the expected molecular weight of a 50S particle without the stalk complex (50S – RL10[RL7/12]$_4$). This highly flexible structure consists of one copy of the RL10 protein bound by two heterodimers of RL12 and RL7. The stalk complex, mediating the interaction of the ribosome with GTPase elongation factors, is only partially annotated in the available high-resolution structural models and the RL7/RL12 proteins are the only ribosomal proteins that do not directly interact with the rRNA[27]. Dissociation of the stalk complex pentamer from the 50S ribosome has been observed before in MS experiments and may be more significant under low magnesium conditions[21,28]. We also detected a less abundant form of the 50S particles with a mass of 1431.5 ± 0.65 kDa, which fits best to 50S particles missing only a single RL7/RL12 heterodimer (50S – [RL7/12]$_2$).

Taken together, our high-resolution native MS data clearly exposes the heterogeneity in the 30S and 50S particles. The heterogeneity in these sub-particles may partly explain why the mass spectra of the 70S particles presented here are less well-resolved, as they evidently represent the sum of the heterogeneous 30S and 50S particles. To explore the limits of our QE-UHMR mass spectrometer even further we performed native MS analysis of Flock House virus (FHV). FHV is an insect virus with a single stranded RNA (ssRNA) genome encapsidated by a protein shell of 33 nanometers in diameter consisting of 180 copies of the 43.6 kDa capsid protein alpha (Cp$_\alpha$). The segmented genome is formed by the 3107 nucleotide RNA1 that encodes an RNA-dependent RNA polymerase and the 1400 nucleotide RNA2 that encodes the capsid protein[29]. Together, the protein and RNA segments form the virus with a theoretical mass of over 9 MDa. We analyzed the authentic virions of FHV purified after infection of cultured *Drosophila* cells using native MS. Figure 3a shows the resulting mass spectrum, a clear series of sharp and well-resolved charge states is observed around 42,000 *m/z*, a region inaccessible to the EMR[11]. Moreover, with an average peak width at half height of around 70 *m/z*, the mass spectrum allows mass measurement with unparalleled mass resolution. The accurate mass of FHV measured here is 9325.5 ± 0.8 kDa, which agrees well with the expected mass when assuming 180 copies of Cp$_\alpha$ and the two RNA segments (9307.2 kDa). Strikingly, top-down tandem MS experiments of the virions revealed the release of two proteins with masses of 39193.23 ± 1.65 Da and 4394.61 ± 0.5 Da (Figure 3b and Supplementary Fig. 4). These masses correspond well with the masses expected for capsid protein beta and the membrane-permeabilizing gamma peptide, formed through auto-proteolytic cleavage of Cp$_\alpha$ during viral maturation[30,31]. Since the native mass spectrum of the intact virion appears to represent a relatively homogeneous population of virus particles, and no Cp$_\alpha$ could be detected, we conclude not only that the auto-proteolytic cleavage is complete but also that majority if not all peptide gamma copies remain non-covalently associated to the capsid after they have been formed.

Here, we have described a new QE-UHMR mass spectrometer with a substantially extended mass range and increased sensitivity and mass resolution at high *m/z*. This new instrument allows high-fidelity, hypothesis-free mass analysis of *E. coli* ribosome particles revealing the sub-stoichiometric association of the elusive small protein *sra*.

The exposed heterogeneity in the particles should enable native MS of ribosomal particles from both prokaryotic and eukaryotic sources as a quality control step prior to subsequent analysis by electron microscopy or crystallography. Furthermore, the well-resolved spectrum of the nearly 10 MDa Flock House virus exposed the fate of the membrane lytic gamma peptide. This type of high-resolution native MS analysis could be extended to study ribosomal binding of drugs, antibiotics and/or initiation/elongation factors, detection of low copy number viral proteins such as polymerases or proteases, as well as to the analysis of numerous other complex macromolecular machines. By improving the accessibility to this type of mass spectrometry research we foresee a significant increase in its application in future hybrid structural biology studies.

Acknowledgements

M.W., K.L.F. and A.J.R.H. are funded by The Netherlands Organization for Scientific Research (NWO) supported large-scale proteomics facility Proteins@Work (Project 184.032.201) embedded in The Netherlands Proteomics Centre. M.W. and A.J.R.H. are additionally supported by a Projectruimte grant (12PR3303-2) from Fundamenteel Onderzoek der Materie (FOM). A.M. and A.J.R.H. acknowledge additional support through the European Union Horizon 2020 programme FET-OPEN project MSmed, Project 686547. A.R. is supported by a University of Texas Medical Branch (UTMB) startup fund and the Texas Rising STARs Award from the University of Texas System. The authors would like to thank Philip Lössl and Fan Liu (Utrecht University) for assistance in the bottom-up LC-MS/MS analysis.

Contributions

M.W., K.L.F., D.B., and M.R. performed the experiments. M.W., K.L.F., A.M. and A.J.R.H. wrote the manuscript. A.R. provided Flock House virus material. A.M. and A.J.R.H. supervised the modifications on the Orbitrap mass analyzer. A.M. and A.J.R.H. conceptually designed the work.

Competing financial interests

D.B., M.R. and A.M. are employees of Thermo Fisher Scientific, the manufacturer and supplier of Orbitrap based mass spectrometers.

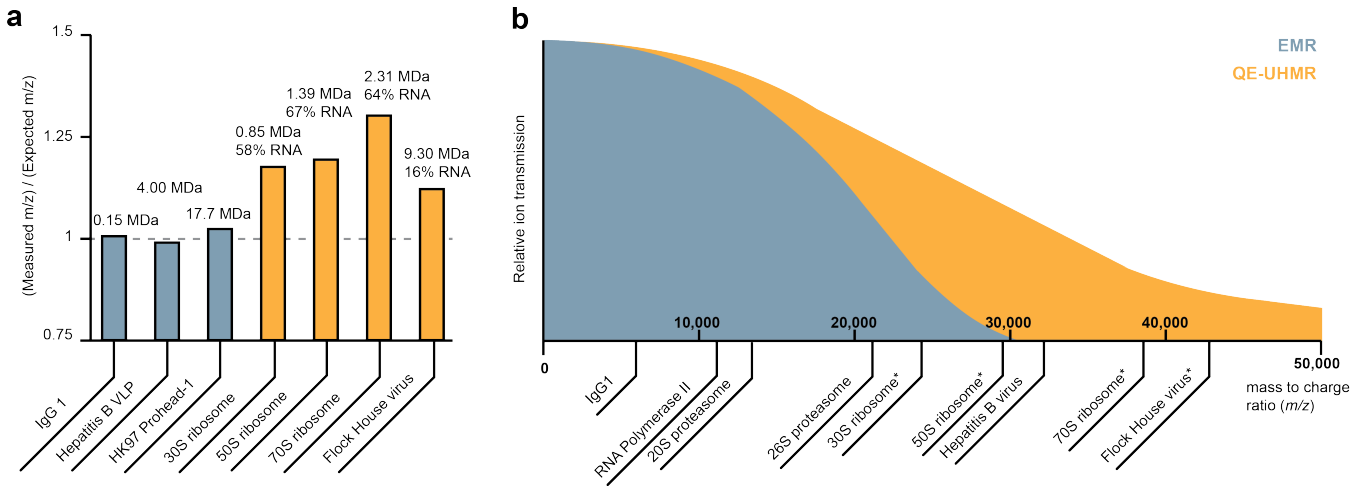
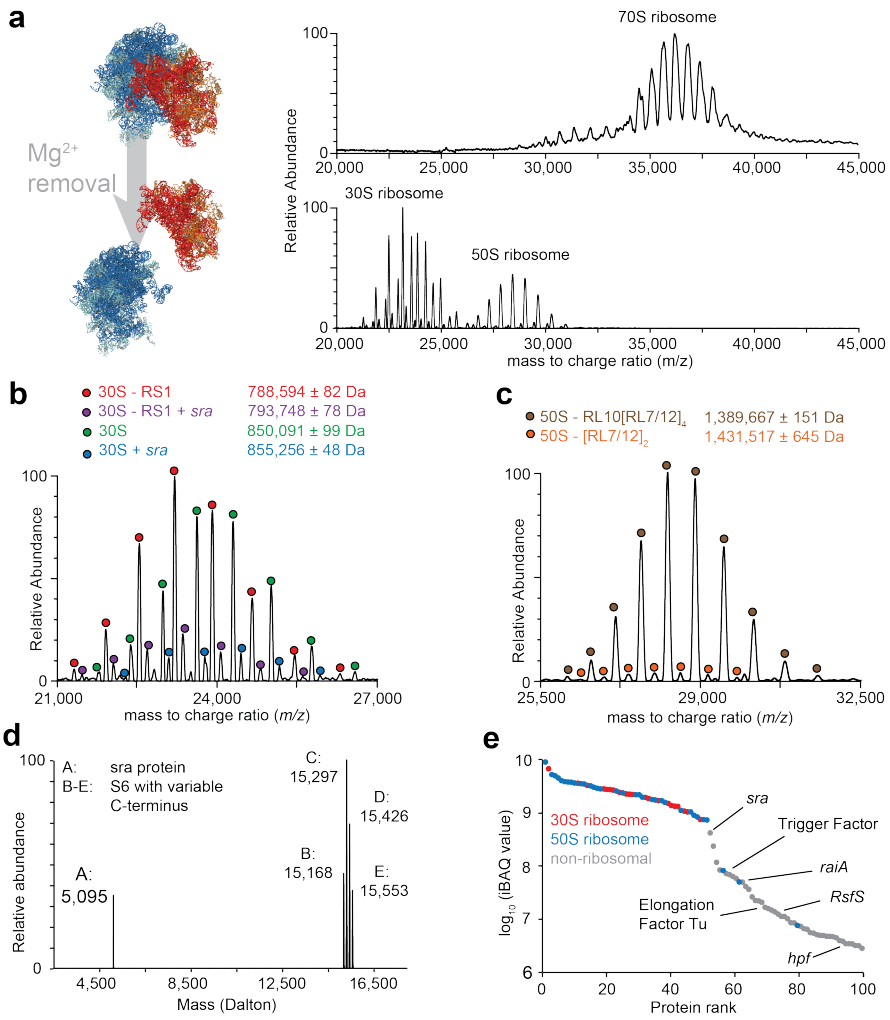


Figure 1 | Increased sensitivity for high *m/z* ions enables high-resolution mass analysis of protein–nucleic acid macromolecular machineries. **a**. Ratio of the measured and expected *m/z* value for a number of protein assemblies (blue) and protein-nucleic acid assemblies (orange, RNA content as a percentage of the molecular weight). The values for protein-only assemblies are taken from literature. **b.** Ion transmission efficiency plotted against *m/z* values for the QE-UHMR (orange) and the EMR (blue). The average *m/z* values of several protein (-nucleic acid) complexes have been annotated, assemblies measured in this work are labeled with an asterisk.

Figure 2 | Accurate mass measurement of 70S, 50S and 30S ribosome particles. **(a)** Left: structures of the *E. coli* 70S ribosome consisting of the 50S (proteins in blue and rRNA in cyan) and 30S (proteins in red and rRNA in orange) particles. Right: native mass spectra of the 70S, 30S and 50S ribosomes. Charge states of all these particles were well-resolved even at 36,000 *m/z*. **(b)** Distinct particles of the 30S ribosome. Masses are shown as mean ± s.d. **(c)** Distinct particles of the 50S ribosome. In most of the particles the pentameric stalk complex is absent. **(d)** Q-ToF tandem MS spectra reveal the release of the *sra* protein and the RS6 protein that has a variable number of glutamate residues at the C-terminus. **(e)** Abundance of proteins in the ribosome preparation. The *sra* protein is substantially more abundant than other stress related proteins (*raiA*, *RsfS* and *hpf*) or common ribosome interacting proteins (see also Supplemental table 5).



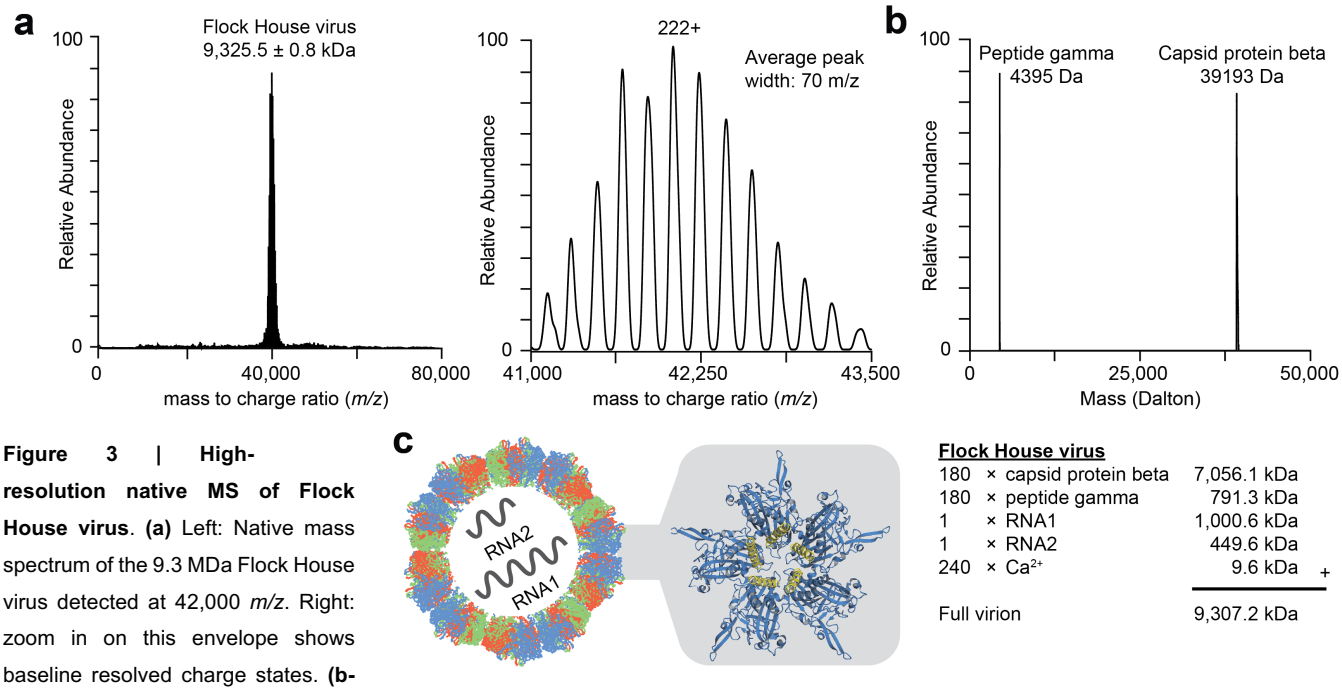


Figure 3 | High-resolution native MS of Flock House virus. **(a)** Left: Native mass spectrum of the 9.3 MDa Flock House virus detected at 42,000 *m/z*. Right: zoom in on this envelope shows baseline resolved charge states. **(b-c)** Q-ToF tandem MS of the FHV particles reveals that the capsid protein alpha in the particles is fully cleaved into the capsid protein beta and the gamma peptide. Structure of the FHV 5-fold symmetry axis (PDB entry 4FTE) shows the capsid protein beta (blue) and the cleaved peptide gamma (yellow).

References

1. Greber, B. J. *et al.* Ribosome. The complete structure of the 55S mammalian mitochondrial ribosome. *Science (New York, N.Y.)* **348,** 303–308 (2015).
2. Plaschka, C. *et al.* Architecture of the RNA polymerase II-Mediator core initiation complex. *Nature* **518,** 376–380 (2015).
3. Bui, K. H. *et al.* Integrated structural analysis of the human nuclear pore complex scaffold. *Cell* **155,** 1233–1243 (2013).
4. Wang, G. *et al.* Molecular Basis of Assembly and Activation of Complement Component C1 in Complex with Immunoglobulin G1 and Antigen. *Molecular Cell*.
5. Diebolder, C. A. *et al.* Complement is activated by IgG hexamers assembled at the cell surface. *Science (New York, N.Y.)* **343,** 1260–1263 (2014).
6. Zhang, X. *et al.* In situ structures of the segmented genome and RNA polymerase complex inside a dsRNA virus. *Nature* **527,** 531–534 (2015).
7. Heck, A. J. Native mass spectrometry: a bridge between interactomics and structural biology. *Nature methods* **5,** 927–933 (2008).
8. Jore, M. M. *et al.* Structural basis for CRISPR RNA-guided DNA recognition by Cascade. *Nature structural & molecular biology* **18,** 529–536 (2011).
9. Zhou, M. *et al.* Mass spectrometry of intact V-type ATPases reveals bound lipids and the effects of nucleotide binding. *Science (New York, N.Y.)* **334,** 380–385 (2011).
10. Sharon, M. & Robinson, C. V. The role of mass spectrometry in structure elucidation of dynamic protein complexes. *Annual review of biochemistry* **76,** 167–193 (2007).
11. Rose, R. J., Damoc, E., Denisov, E., Makarov, A. & Heck, A. J. High-sensitivity Orbitrap mass analysis of intact macromolecular assemblies. *Nature methods* **9,** 1084–1086 (2012).
12. Gault, J. *et al.* High-resolution mass spectrometry of small molecules bound to membrane proteins. *Nature methods* **13,** 333–336 (2016).
13. van de Waterbeemd, M. *et al.* Simultaneous assessment of kinetic, site-specific, and structural aspects of enzymatic protein phosphorylation. *Angewandte Chemie (International ed. in English)* **53,** 9660–9664 (2014).
14. Snijder, J. *et al.* Defining the stoichiometry and cargo load of viral and bacterial nanoparticles by Orbitrap mass spectrometry. *Journal of the American Chemical Society* **136,** 7295–7299 (2014).
15. van de Waterbeemd, M. *et al.* Examining the Heterogeneous Genome Content of Multipartite Viruses BMV and CCMV by Native Mass Spectrometry. *Journal of the American Society for Mass Spectrometry* **27,** 1000–1009 (2016).
16. Belov, M. E. *et al.* From protein complexes to subunit backbone fragments: a multi-stage approach to native mass spectrometry. *Analytical chemistry* **85,** 11163–11173 (2013).

17. Wilson, D. N. & Nierhaus, K. H. Ribosomal proteins in the spotlight. *Critical reviews in biochemistry and molecular biology* **40,** 243–267 (2005).

18. Arnold, R. J. & Reilly, J. P. Observation of Escherichia coli ribosomal proteins and their posttranslational modifications by mass spectrometry. *Analytical biochemistry* **269,** 105–112 (1999).

19. Decatur, W. A. & Fournier, M. J. rRNA modifications and ribosome function. *Trends in biochemical sciences* **27,** 344–351 (2002).

20. Rostom, A. A. *et al.* Detection and selective dissociation of intact ribosomes in a mass spectrometer. *Proceedings of the National Academy of Sciences of the United States of America* **97,** 5185–5190 (2000).

21. Deroo, S. *et al.* Mechanism and rates of exchange of L7/L12 between ribosomes and the effects of binding EF-G. *ACS chemical biology* **7,** 1120–1127 (2012).

22. Laszlo, K. J. & Bush, M. F. Analysis of Native-Like Proteins and Protein Complexes Using Cation to Anion Proton Transfer Reactions (CAPTR). *Journal of the American Society for Mass Spectrometry* **26,** 2152–2161 (2015).

23. Lauber, M. A., Rappsilber, J. & Reilly, J. P. Dynamics of ribosomal protein S1 on a bacterial ribosome with cross-linking and mass spectrometry. *Molecular & cellular proteomics : MCP* **11,** 1965–1976 (2012).

24. Sorensen, M. A., Fricke, J. & Pedersen, S. Ribosomal protein S1 is required for translation of most, if not all, natural mRNAs in Escherichia coli in vivo. *Journal of molecular biology* **280,** 561–569 (1998).

25. Shine, J. & Dalgarno, L. The 3'-terminal sequence of Escherichia coli 16S ribosomal RNA: complementarity to nonsense triplets and ribosome binding sites. *Proceedings of the National Academy of Sciences of the United States of America* **71,** 1342–1346 (1974).

26. Izutsu, K. *et al.* Escherichia coli ribosome-associated protein SRA, whose copy number increases during stationary phase. *Journal of bacteriology* **183,** 2765–2773 (2001).

27. Diaconu, M. *et al.* Structural basis for the function of the ribosomal L7/12 stalk in factor binding and GTPase activation. *Cell* **121,** 991–1004 (2005).

28. Ilag, L. L. *et al.* Heptameric (L12)6/L10 rather than canonical pentameric complexes are found by tandem MS of intact ribosomes from thermophilic bacteria. *Proceedings of the National Academy of Sciences of the United States of America* **102,** 8192–8197 (2005).

29. Odegard, A., Banerjee, M. & Johnson, J. E. Flock house virus: a model system for understanding non-enveloped virus entry and membrane penetration. *Current topics in microbiology and immunology* **343,** 1–22 (2010).

30. Fisher, A. J. & Johnson, J. E. Ordered duplex RNA controls capsid architecture in an icosahedral animal virus. *Nature* **361,** 176–179 (1993).

31. Bothner, B., Dong, X. F., Bibbs, L., Johnson, J. E. & Siuzdak, G. Evidence of viral capsid dynamics using limited proteolysis and mass spectrometry. *The Journal of biological chemistry* **273,** 673–676 (1998).