## ****Interpretation of the Rat Viral Capsid Multiple Sequence Alignment (Clustal Omega)****

### ****What the Alignment Represents****

The file contains multiple **capsid protein sequences** from **rat-associated viruses** and a **reference structure (6WFT)**.  
All of these proteins are **related**, forming the protective shell (capsid) around the virus’s RNA.  
The alignment shows **how similar or different** each amino acid position is across these viruses.

Each line represents part of the protein sequence, and the alignment marks help interpret conservation:

1. \* → All residues are identical (fully conserved)
2. : → Highly similar residues (same type/charge)
3. . → Weakly similar residues
4. (no mark) → Variable region

### ****Overall Summary of Similarity****

1. **Average identity:** ~85–90% (very high — shows they are from the same viral family, possibly Hepevirus).
2. **Conserved residues:** Many clusters of \* appear through the alignment, especially in the middle and near the end.
3. **Variable regions:** Found mostly at the **surface-exposed loops**, shown by gaps (---) and changes in letters.

This pattern is typical for **capsid proteins**, which must keep their **core structure stable** but allow **surface variation** for immune evasion and host adaptation.

### ****Conserved Regions****

| **Approx. Region** | **Sequence Example** | **Conservation** | **Likely Function** |
| --- | --- | --- | --- |
| **1–60** | MSFVDHPPDWLEKIGEGFKEFLNLEPGPPKPKPGYQDNARGLVVPGYKYLGPFNGLD | Highly conserved | Initiates folding; forms N-terminal scaffold for capsid formation. |
| **220–300** | GDVGQSSGNWHCDSVWMGDRVL...TKSTRTWSLPTYNNHLYKQIN | Very conserved core | Contains the β-barrel fold typical of capsid proteins. Maintains 3D shell stability. |
| **350–420** | HWGIRPRRLNFKLFNIQVKEVTTTDGTKTIANNLTSTVQVFADTEHQLPYILGSAHEGCM | Fully conserved | Rich in charged residues (R, K, H) that bind viral RNA and strengthen the capsid interior. |
| **650–715** | WQDRDIYLQGPIWAKIPETDGHFHPSPLMGGFGLKNPPPQILIKNTPVPADPPTQ | Highly conserved | Flexible hinge region; allows subunits to curve and close the viral shell. |
| **720–730** | IGTRYLTHTL | 100% conserved | C-terminal closure; final locking segment that stabilizes the capsid. |

**Interpretation:**  
These conserved regions act like the **steel framework** of the virus.  
They ensure the capsid has the correct shape, rigidity, and ability to pack RNA efficiently.

### ****Variable Regions****

Found mainly between **residues 100–200** and **450–600**, where you see alignment gaps or mismatched residues.These are likely **surface-exposed loops** that can tolerate mutations.  
They determine **how the virus interacts with host cells** and **how visible it is to the immune system.**

🟡 **Function:**

* Adapt to different rat species or host receptors
* Avoid recognition by antibodies
* Maintain inectivity while changing appearance

### 🔹 ****5. Conserved Motifs and Their Biological Roles****

| **Motif** | **Position** | **Function** | **Impact** |
| --- | --- | --- | --- |
| **GDVGQSSGNWHC** | ~230–240 | Core fold motif (β-barrel loop) | Keeps capsid structurally intact. |
| **TKSTRTWSLPTYNNHLYKQIN** | ~260–280 | Assembly motif | Helps capsid subunits fit together. |
| **HWGIRPRRLNFKLFNIQVKEVTTTDGTKTIANNL** | ~350–380 | RNA-binding motif | Positively charged (R, K) residues interact with viral RNA. |
| **PPPQ / PVP / PPT motifs** | ~640–660 | Proline-rich hinge | Adds flexibility for curvature during assembly. |
| **IGTRYLTHTL** | ~720–730 | Closing motif | Locks the capsid and stabilizes final structure. |

**Interpretation:**  
Each motif contributes a specific mechanical or chemical role:

* Glycine (G) → flexibility
* Proline (P) → bending points
* Arginine (R) / Lysine (K) → positive charge → RNA binding
* Cysteine (C) → disulfide bonds → extra stability