

Bangladesh Proposal for UTC-162

**Proposal for changing the Composition-Exclusions of Bangla Characters (Bengali)**

Submitted by Mohammad Mamun Or Rashid (on behalf of ICT Division, Bangladesh)

**Proposal:** Request to change Bangla three atomic character RRA ড (09DC), RHA ঢ (09DD) and YYA য় (09DF) from the composition-exclusion list.

**Rationale:** Initial consideration for these three characters would be atomic, then decomposed. Enstilstment of their atomic code-points in composition-exclusion is controversial.

Bangla (Bengali) Composition-Exclusions<sup>1</sup> in UNICODE NORMALIZATION FORMS, only decomposed form for some Bangla ‘additional characters’ (RRA, RHA, YYA) are allowed while their corresponding atomic code-point is blocked. It is assumed that atomic code-point is exposed to code-chart as it is a primary representation and will be protected by Unicode stability principle. Our proposal is significant, allowing atomic code-point only and disallow the corresponding decomposed one.

Character	Atomic	Decomposed
RRA ড	09DC	09A1+09BC= 09DC
RHA ঢ	09DD	09A2+09BC=09DD
YYA য়	09DF	09AF+09BC=09DF

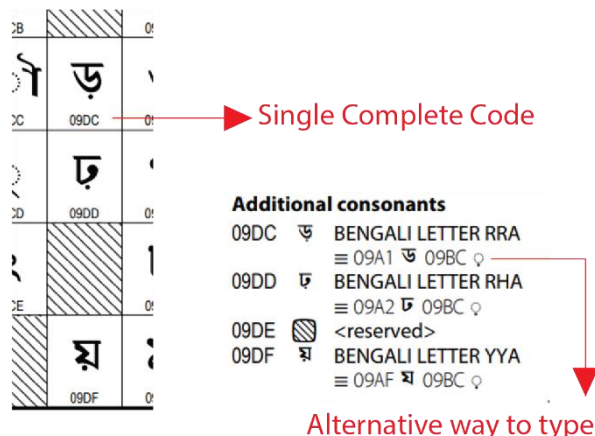


Figure 1. These characters have two code-points where atomic one should not be blocked.

The primary reason behind this occurrence is these three characters (ড,ঢ,য়) are not in the other Indic-script (especially Devanagari/ ISCII standard) or other New-Brahmi script code-chart. So other Indic-script users are habituated with compound form. However, Bengali speakers never do that as these characters are their primary character.

This complexity will create data anomalies, especially in Bangla corpus and thus use of Bangla language in the digital platform will be disrupted.

Also, it created another discrepancy at IDNA2008 protocol, an application based on Unicode standard. IDNA2008 simply reflects the definition of NFC:

<sup>1</sup> <http://www.unicode.org/Public/12.0.0/ucd/CompositionExclusions.txt>

Inconsistency in IDNA2008 protocol

```
09DC..09DD : DISALLOWED # BENGALI LETTER RRA..BENGALI LETTER RHA
09DF       : DISALLOWED # BENGALI LETTER YYA
09BC..09C4 : PVALID      # BENGALI SIGN NUKTA..BENGALI VOWEL SIGN VOCAL
0B5F..0B63 : PVALID      # ORIYA LETTER YYA..ORIYA VOWEL SIGN VOCALIC L
0A5C       : PVALID      # GURMUKHI LETTER RRA
```

Therefore, to reduce the discrepancy in Bangla Corpus and inconsistency in the IDNA protocol, we would like to propose to consider our proposal.

**Submitted by** Mohammad Mamun Or Rashid (on behalf of ICT Division, Bangladesh)  
Bangla language Technology Specialist, Enhancement of Bangla Language in ICT through Research and Development Project, Bangladesh Computer Council, ICT Division, Government of Bangladesh.  
& Assistant Professor, Jahangirnagar University, Bangladesh

Email: [mamunbd@juniv.edu](mailto:mamunbd@juniv.edu), [mamunbd@bcc.gov.bd](mailto:mamunbd@bcc.gov.bd)