Doc Type:	Working Group Document
Title:	Feedback on IRGN2578 “Comments on encoding early Chinese organic chemical character in WS2021 and other complex ideographs”[1]
Source:	Huáng Jùnliàng (黃俊亮)
Status:	Individual contribution
Action required:	To be considered by the IRG and UTC
Date:	March 12, 2023

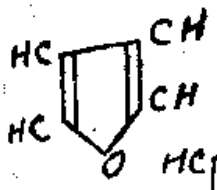
In IRGN2578, Eiso Chan listed four base characters of the first type of early organic chemical characters: 烷, 烯, 炔 and 圈. Additionally, Chan noted the importance of checking historical usage of numerals greater than 10 when combined with these base characters. To recap, the combined characters are instances of the hydrocarbon nomenclature first proposed by 梁國常 in his work 有機化學命名芻議 [1].

Table 1 contains a list of unencoded hydrocarbon characters based on evidence from Liang's work.

Table 1: Unencoded hydrocarbon characters in Liang's work not submitted for WS2021

Character	Page	Evidence
烷	73	Saturated hydrocarbons C_nH_{2n+2} 烷(音充)
烯	73	Olefines series C_nH_{2n} 烯(音欠)
炔	73	Acetylene series C_nH_{2n-2} 炔(音少)
圈	73	Benzene ring and Furane ring 圈(音圈)
烷	73	Propane 烷(讀充三)
歎	78	3-chloro-1-propene $CH_2:CH\cdot CH_2Cl$ 歎[1] 氯[3]
炔	78	allylene C_3H_6 炔

[1]Sources of this document are available online: `https://github.com/JLHwung/IRGN2578Feedback`.

Character	Page	Evidence
𪜀	73	Butane 𪜀 (讀 充 圍)
𪜁	78	butylene C_4H_8 𪜁
𪜂	81	pentamethylene diamine $NH_2(CH_2)_5NH_2$ 𪜂 二 胺[1:5]
𪜃	73	Furane 𪜃 (讀 圍 五)
		furane  𪜃 氧
		thiophene 𪜃 硫
		pyrrole 𪜃 𪜄
	88	
𪜅[2]	81	glyceryl tripalmitate (tripalmitin) $C_3H_5(O{\cdot}COC_{15}H_{31})_3$ 三 𪜅 酸 𪜅
	81	monopalmitin $(HO)_2C_3H_5OCOC_{15}H_{31}$ 𪜅 酸 𪜅 二 酯
𪜆	80	monde cylic acid $C_{18}H_{37}COOH$ 𪜆 酸
𪜇	78	heneicosane $C_{21}H_{44}$ 𪜇
𪜈	78	hexacontane $C_{60}H_{122}$ 𪜈

[2]To clarify, in the presented evidences, it would be more accurate to use 𪜅 instead of 𪜅, as both tripalmitin and monopalmitin contain hexadecanoate ($C_{16}H_{31}O_2^-$) rather than pentadecanoate ($C_{15}H_{29}O_2^-$).

If we are to encode 氿 (00016), 氿 (00017), 歐 (01900), 氽 (00014) and 圈 (00777) one by one, then we should also encode the other 15 characters above in order to digitize Liang's work. In total, we need to encode 20 characters, which is already more than the length of a minimal Unicode block (16 characters), should they be encoded as IDJ sequences proposed in IRGN2578.

Additionally, it's important to note that hydrocarbons can have an arbitrary number of carbon atoms, making such characters an open set. It's possible that Liang's nomenclature was adopted in other historical documents, which may contain additional hydrocarbon characters that we're currently unaware of. In other words, we may have to encode more than 20 hydrocarbon characters in the future.

To summarize, encoding hydrocarbon characters as IDJ sequences can help us save codepoint space and address the open set nature of Liang's nomenclature. As presented in IRGN2578, the IDJ approach potentially solve other encoding issues, too. In light of this I recommend that we carefully consider Chan's approach and inform our encoding strategy.

Acknowledgements

"Hulenkius" helps me polish the font.

References

[1] 梁國常, "有機化學命名芻議," 北京大學月刊, vol. 7, pp. 71–89, 1920. [Online]. Available: http://read.nlc.cn/allSearch/searchDetail?searchType=all&showType=1&indexName=data_404&fid=01J000340

(End of Document)