**Appendix 1.** The protein sequences used as the Q-string benchmark – Q16778.

|  |  |
| --- | --- |
| Position | Sequence |
| 6 | MPEPAKSAPAPKKGSKKA |
| 12 | MPEPAKSAPAPKKGSKKAVTKAQK |
| 13 | MPEPAKSAPAPKKGSKKAVTKAQKK |
| 16 | EPAKSAPAPKKGSKKAVTKAQKKDGKK |
| 17 | PAKSAPAPKKGSKKAVTKAQKKDGKKR |
| 21 | APAPKKGSKKAVTKAQKKDGKKRKRSR |
| 24 | PKKGSKKAVTKAQKKDGKKRKRSRKES |
| 25 | KKGSKKAVTKAQKKDGKKRKRSRKESY |
| 28 | SKKAVTKAQKKDGKKRKRSRKESYSIY |
| 29 | KKAVTKAQKKDGKKRKRSRKESYSIYV |
| 31 | AVTKAQKKDGKKRKRSRKESYSIYVYK |
| 35 | AQKKDGKKRKRSRKESYSIYVYKVLKQ |
| 44 | KRSRKESYSIYVYKVLKQVHPDTGISS |
| 47 | RKESYSIYVYKVLKQVHPDTGISSKAM |
| 58 | VLKQVHPDTGISSKAMGIMNSFVNDIF |
| 86 | RIAGEASRLAHYNKRSTITSREIQTAV |
| 109 | QTAVRLLLPGELAKHAVSEGTKAVTKY |
| 117 | PGELAKHAVSEGTKAVTKYTSSK |
| 121 | AKHAVSEGTKAVTKYTSSK |
| 126 | SEGTKAVTKYTSSK |