2 | METHODS The Biosaur performance has been evaluated using publicly available (ProteomeXchange dataset PXD010012) timsTOF data.7 The data were obtained for a 200 ng HeLa mammalian protein digest sample using a 120 min LC gradient and a timsTOF Pro mass spectrometer operated in parallel accumulation–serial fragmentation (PASEF) mode. Negative ion mode tests were performed for data obtained using ultrashort 5 min LC gradients (Dionex UltiMate 3000 HPLC system, Thermo Fisher Scientific, Waltham, MA, USA) and a high-resolution Orbitrap FTMS mass spectrometer (Q-Exactive HF, Thermo Fisher Scientific). MS1 spectra of 200 ng of HeLa were acquired at a mass resolving power of 60 000 at m/z 200 in both positive and negative ion modes. Details of the experimental setup are described elsewhere.23 Here 5 min LC/MS1 and LC/MS/MS data obtained for HeLa in the latter study were used for in-source fragmentation and semi-tryptic peptide searches, as well as for the Biosaur performance comparison with other feature-detection alternatives. The IdentiPy search engine24 with default settings and an enabled semi-tryptic peptide search option was used. Biosaur was employed for MS1 feature detection. 3 | RESULTS 3.1 | Software design The principle of the Biosaur algorithm is shown schematically in Figure 1. It takes the input files in mzML format and provides the output files with detected peptide features in \*.tsv (tab-separated values) format. Contrary to the many feature-detection algorithms that process peptide MS1 spectra one-by-one, Biosaur is based on the so-called “hill” concept introduced by Teleman et al.22 The algorithm's work starts with combining centroid ion peaks into hills, which are the groups of peaks with similar m/z values in consecutive MS1 scans. All peaks in the mass spectra from the first scan are converted into the constructed hills. After that, all peaks from the second scan are added to the current hills under construction if the difference between the hill's m/z value and the added peak's m/z value is within the user-defined mass measurement accuracy (MMA). The MMA is 8 ppm by default. The hill's m/z value is calculated as the average of the m/z values of the hill's peaks from the last three scans. If several peaks in the subsequent scan have m/z values close to that of one of the hills under construction, only the peak with the smallest mass difference is added to that hill. All peaks with m/z values far (by MMA) from the hills under construction are used to start new hills. After peak matching, the Biosaur finds hills that were not updated with the new peaks for the two previous scans. These hills are then added to the group of finished hills and do not participate further in the next peak-matching steps. The above procedure continues through all scans until the end of the run file. Upon reaching the last scan, all hills under construction are converted into the final group of finished hills. Finally, all finished hills with scan lengths less than the user-defined minimum (three scans by default) are discarded. As a result, the Biosaur collects a list of finished hills with the corresponding information about the peaks added to each of them, including m/z values, peak intensities, and scan identifications. It also calculates the hill m/z values as the weighted by intensity average of the m/z values of the peaks forming a particular hill. The hills are gathered into 13C isotope clusters in the next stage of the algorithm's workflow. Biosaur detects potential isotopic clusters using cosine similarities between the RT profiles of monoisotopic and isotopic hills, cluster peak abundances, masses, and the averaging model.25 First, Biosaur goes through all the hills and determines those which have similar RTs and a mass shift equal to N × 1.00335/charge, where N is the 13C isotope number from 1 up to 10 and charge is a potential charge from a user-defined range. User-defined thresholds for the minimal RT profile cosine correlation and mass accuracy are used. Biosaur does not check the N + 1 isotope if the N isotope was not found. The similarity of isotopic patterns is calculated using cosine correlation between theoretical and experimental intensity profiles. The averaging model is used for calculation of the peptide isotopes and the corresponding isotopic distributions. The 13C isotopic distribution is calculated by assuming that every 111.1254 Da contains 4.9384 carbon atoms. To reduce the number of false positive matches, cosine similarities are evaluated by filling zero values for the non-overlapping 2 of 8 ABDRAKHIMOV ET AL. scans between the hills. Intensity-weighted m/z values across all scans are reported as the monoisotopic m/z of the finally determined peptide features. In addition, Biosaur reports intermediate metrics, such as the m/z values for all isotopic peaks, coefficients of variations between scans for each of these m/z values, experimental isotopic intensity distributions, etc. These values can be later used as the metrics of detected feature quality. For example, one can more accurately recalculate correlation between theoretical and experimental isotopic intensity distributions during the following post-search analysis of the matching between peptide sequences and the feature. Note that an accurate theoretical profile can be used instead of the averaging approximation in that latter step. At the third stage of Biosaur's workflow, all potential isotopic clusters are sorted in decreasing order by the sum of the numbers of visible isotopes and cosine correlation for the corresponding average molecule (Figure 2). A list of all detected isotopic clusters is generated one-by-one and all hills that contributed to the already added clusters are removed from further consideration. The key functionalities of the Biosaur algorithm are presented in Table 1. Biosaur has a targeted mode, in which it matches the results of identification of MS/MS spectra to the peptide features. The current Biosaur version also supports X!Tandem,26 IdentiPy,24 MSFragger,27 and Comet28 search engine outputs in pepXML format, as well as MSGF+29 output in mzID format.