Corrigendum: Exploring the sequence determinants of amyloid structure using position-specific scoring matrices

Sebastian Maurer-Stroh, Maja Debulpaep, Nico Kuemmerer, Manuela Lopez de la Paz, Ivo Cristiano Martins, Joke Reumers, Kyle L Morris, Alastair Copland, Louise Serpell, Luis Serrano, Joost W H Schymkowitz & Frederic Rousseau
Nat. Methods **7, 237–242 (2010); published online 14 Feburary 2010; corrected after print 29 September 2010.**

In the version of this paper originally published, the name of and reference to the algorithm in the rightmost column of Table 1 were incorrect. The correct reference (ref. 40) has been added in the paper. The error has been corrected in the PDF and HTML versions of the article.

Addendum: Exploring the sequence determinants of amyloid structure using position-specific scoring matrices

Sebastian Maurer-Stroh, Maja Debulpaep, Nico Kuemmerer, Manuela Lopez de la Paz, Ivo Cristiano Martins, Joke Reumers, Kyle L Morris, Alastair Copland, Louise Serpell, Luis Serrano, Joost W H Schymkowitz & Frederic Rousseau
Nat. Methods **7, 237–242 (2010); published online 14 Feburary 2010; addendum published after print 29 September 2010.**

After the publication of our paper, we identified a mistake in Table 1 regarding the comparison of our program, Waltz, to the program 3D profile[1] (ref. 25 in our paper); we cited the wrong name and reference of the algorithm in the right column. This error has been corrected after print to refer to the algorithm we actually used, the method described in reference 2 (ref. 40 in the corrected paper). However, as the 3D profile[1] method developed in the Eisenberg laboratory has a long-standing good reputation as an amyloid prediction tool, here we compare it to Waltz. An improved version of 3D profile[3] was published about a week and a half before our paper, so for complete transparency we also compare Waltz to the improved 3D profile algorithm.

In **Table 1** we list all predicted peptides and scores or energies, respectively, comparing Waltz (threshold 77, running on our web-server at http://waltz.switchlab.org/) with the 3D profile[1] scores at the ZipperDB website (http://services.mbi.ucla.edu/zipperdb; energy threshold was –23; additional shape complementarity > 0.7 for the 3D profile 2010 version[3]). The sensitivity of 3D profile on our sup35 positive set was 67% (75% if one includes prediction of a hexapeptide that is almost but not fully included in the tested decapeptide).

However, the higher sensitivity of 3D profile comes at a cost of lower specificity (more false positives). To estimate the rate of false positives, we derived a reliable negative set from our experimental data for sup35, which included all decapeptides that did not form fibers under the unified experimental conditions and did not overlap with any positively tested one (31 in total). However, we cannot draw hard conclusions as the availability of bona fide experimental data is typically limiting and these numbers are too low for a good general comparison. An additional complication is that 3D profile is designed to predict hexapeptides; as next best approximation we defined the best score or energy of a fully included hexapeptide as prediction for the respective peptides. Owing to this limitation and the fact that well-predicted hexapeptides may actually form amyloid fibers and the longer decapeptide does not, it may be wiser to exclude such peptides in an alternative comparison with only 26 'negative' peptides, the reduced benchmark set ('–5') (**Table 1**).

Sensitivities of predictors should either be compared at similar levels of specificity (as should be done in consensus methods, such as AmylPred[4]), or one needs to consider both sensitivity and specificity together. Established measures for this are the Matthew correlation coefficient and the probability excess[5]. Probability excess has the additional advantage that it is also independent of set size inequalities[6], which are not considered in other measures such as accuracy and precision.

The resulting performance statistics are reported in **Table 2**. Although 3D profile 2006 version[1] predicted several additional false positives compared to Waltz, the improved 3D profile 2010 version[3] filtered out several of these. Considering the possibility that high-scoring hexapeptides may indeed form fibers outside of the experimentally tested decapeptide context, the performances of Waltz and 3D profile (2010 version)[3] become comparable over the reduced benchmark set ('–5'). In fact, the observed differences

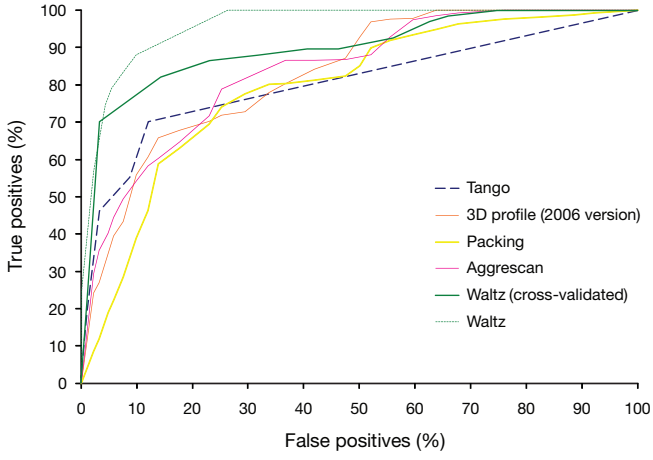

Figure 1 | Comparison of ROC curve performance on the AmylHex dataset.