GraphEnt: a maximum-entropy program with graphics capabilities
Nicholas M. Glykos and Michael Kokkinidis
A maximum-entropy formalism aimed at the production of a ‘maximally noncommittal’ map is a standard method in fields of science like radio-astronomy, but a rare exception in both X-ray crystallography and electron microscopy (or crystallography). This is rather unfortunate, given the wealth of information that a maximum-entropy map can reveal, especially when the map itself is the end product (for example, low-resolution electron or potential density maps, Patterson functions, deformation maps). The program GraphEnt attempts to automate the procedure of calculating maximum-entropy maps, with emphasis on the calculation of difference Patterson functions for macromolecular crystallographic problems, while providing a useful graphical output of the current stage of the calculation.
Typical run times: these depend greatly on the size of the map, the type of calculation and the quality of the input data. For example, a 262,144 (= 128 × 64 × 32) pixels $mF_o-ex$-[raw]-map corresponding to a reasonably accurate (by macromolecular standards) 3.8 Å data set was calculated in less than 8 min of CPU time on a DEC Alpha 1200, while a 524,288 (= 128 × 128 × 32) pixels difference Patterson map for a weakly substituted derivative (which makes the calculation easier) took only 46 s on the same machine. On the other hand, a 2 Å ($2mF_o - DF_c\exp(i\psi)$) synthesis with 3,072,000 (= 160 × 160 × 120) pixels took ~40 min of CPU time.

Number of lines: 5537 for the source code, 3197 for the raw \textit{LATEX} document.

Test status: several difference Patterson functions for three different crystal forms have been calculated, both in projection and in three dimensions. The program has also been tested with a medium-resolution single isomorphous replacement (SIR) phased protein map, and with an 8 Å resolution cryoelectron-microscopic reconstruction of the potential density projection of a large multiprotein complex.

Additional features: GraphEnt uses the PGPLOT graphics library to plot (using contours and/or grayscale representations) a user-defined section from both the conventional and the maximum-entropy maps. The plot of the maximum-entropy map is updated as the calculation proceeds, allowing the user to identify its most persistent features. Fig. 1 shows an image captured from the screen of a workstation performing a GraphEnt calculation. In the case of an isomorphous difference Patterson calculation, the program also draws the corresponding normal probability plot (Howell & Smith, 1992) which can be used to select suspect data points.
computer programs

2.5. Documentation

Extensive documentation is available with the distribution, in the form of a Postscript file and as an HTML version.

3. Applications

*GraphEnt* can recognize and automatically perform several of the most common types of crystallographic syntheses, as discussed in §2.4. Additionally, any type of synthesis that can be reduced to one of these can also be performed, but the reduction step is the responsibility of the user.

As an example of the application of the program, we present results from an anomalous Patterson function calculation using data collected from a horse heart myoglobin crystal. The data were collected with Cu Kα radiation. The anomalous signal comes from the iron atom of heme (with $\Delta F_{\text{Fe,CuKα}} = 3.2 \, \text{e}^{-}$). Table 1 presents statistical information about this data set. To make the example more realistic, we used only data between 20 and 3 Å resolution, and we simulated the presence of outliers in the data by multiplying the amplitude ($F_{\text{ano}}$) and standard uncertainty ($\sigma(F_{\text{ano}})$) of three randomly chosen strong reflections by a factor of 3.0.

A comparison of the Harker sections ($v = 1/2$) from the conventional and *GraphEnt* maps (two uppermost panels in Fig. 2) is rather striking: the presence of outliers in the data has completely wiped out the signal from the conventional map, leaving behind a checkerboard appearance, which is all too familiar to macromolecular crystallographers. In sharp contrast, the *GraphEnt* map resembles more a map calculated with hypothetical error-free data than an anomalous Patterson function calculated with real data.

Figure 2

Comparison of the Harker ($v = 1/2$) sections from two conventional anomalous Patterson functions and a *GraphEnt* anomalous Patterson function of a myoglobin crystal (see text for details). All three Harker sections are contoured with the first (dashed) contour at the mean density, and then at intervals of 0.5 of the r.m.s. deviation of the densities of the whole map. The two insets show the distribution of normalized density [in units of $\langle \rho - \langle \rho \rangle \rangle / \sigma(\rho)$] through the major peaks (of the respective maps) and in a direction parallel to the longest axes of the peaks. This figure was prepared with the programs *PLUTO* and *PLTDEV* from the CCP4 suite of programs (Collaborative Computational Project, Number 4, 1994) and with the program *XMGR*, available via http://plasma-gate.weizmann.ac.il/Xmgr/.
To make the comparison with the GraphEnt map more meaningful, we also present (lowest panel of Fig. 2) the same Harker section from a conventional map calculated after rejection of the three outliers. Although this time the Fe–Fe peak is (reassuringly) the strongest peak in the conventional map as well, the GraphEnt map (which was calculated with the ‘outliers’ included in the data) is still by far superior. The difference in the appearance of the two syntheses is not the result of a uniform reduction of the contrast of the GraphEnt map; as the two insets in Fig. 2 show, the peak of the GraphEnt synthesis stands at approximately 22σ above the mean density of the map, whereas the same peak from the conventional synthesis is only 10σ above the mean.

In summary, the comparison of the conventional and GraphEnt maps illustrates all the advantages of the maximum-entropy formalism that are usually cited in the literature: (i) the maximum-entropy map, by being the most uniform map consistent with the observations, only shows detail for which there is evidence in the data, (ii) the effects arising from the presence of outliers in the data are greatly reduced, (iii) the noise level and side lobes (due to series termination errors) are greatly reduced, (iv) the map is everywhere positive and as smooth as the data allow, and (v) the maximum resolution consistent with the data is achieved.

4. Program availability

The source code of the program, together with its documentation and some example scripts, is distributed free of charge to both academic and non-academic users, and is immediately available for download via http://origin.imbb.forth.gr:8888/~glykos/. The distribution also contains stand-alone executable images suitable for the Silicon Graphics and DEC Alpha OSF workstation architectures.

References


\[^2\] The word ‘outlier’ is used here catachrestically: as long as the standard uncertainties are correctly estimated, there is nothing wrong with the measurements of these reflections. The common macromolecular practice to exclude large differences from the calculation of Patterson functions arises from the inability of the conventional synthesis to deal correctly with incomplete and noisy data.