

Structural bioinformatics

## THESEUS: maximum likelihood superpositioning and analysis of macromolecular structures

Douglas L. Theobald\* and Deborah S. Wuttke

Department of Chemistry and Biochemistry, University of Colorado at Boulder, Boulder, CO 80309-0215, USA

Received on May 2, 2006; accepted on June 12, 2006

Advance Access publication June 15, 2006

Associate Editor: Dmitrij Frishman

### ABSTRACT

**Summary:** THESEUS is a command line program for performing maximum likelihood (ML) superpositions and analysis of macromolecular structures. While conventional superpositioning methods use ordinary least-squares (LS) as the optimization criterion, ML superpositions provide substantially improved accuracy by down-weighting variable structural regions and by correcting for correlations among atoms. ML superpositioning is robust and insensitive to the specific atoms included in the analysis, and thus it does not require subjective pruning of selected variable atomic coordinates. Output includes both likelihood-based and frequentist statistics for accurate evaluation of the adequacy of a superposition and for reliable analysis of structural similarities and differences. THESEUS performs principal components analysis for analyzing the complex correlations found among atoms within a structural ensemble.

**Availability:** ANSI C source code and selected binaries for various computing platforms are available under the GNU open source license from <http://monkshood.colorado.edu/theseus/> or <http://www.theseus3d.org>

**Contact:** douglas.theobald@colorado.edu

**Supplementary Information:** Supplementary data including details of the ML superpositioning algorithm are available at *Bioinformatics* online.

### 1 INTRODUCTION

Superpositioning macromolecular structures is an essential tool in structural bioinformatics and is used routinely in the fields of NMR, X-ray crystallography, protein folding, molecular dynamics, rational drug design and structural evolution (Bourne and Shindyalov, 2003; Flower 1999). Superpositioning allows comparison of structures by fitting their atomic coordinates to each other as closely as possible. The valid interpretation of a superposition relies upon the quality of the estimated orientations of the molecules, and thus reliable and robust superpositioning tools are a critical component of structural analysis and comparison.

The structural superposition problem has classically been solved with the standard statistical optimization method of least-squares (LS) (Flower, 1999). The LS objective is to find the rotations and translations that minimize the squared distances among corresponding atoms in the observed structures. A fundamental justifying assumption of LS (as given in the Gauss–Markov theorem) requires that the errors have equal variance (Seber and Wild, 1989). When

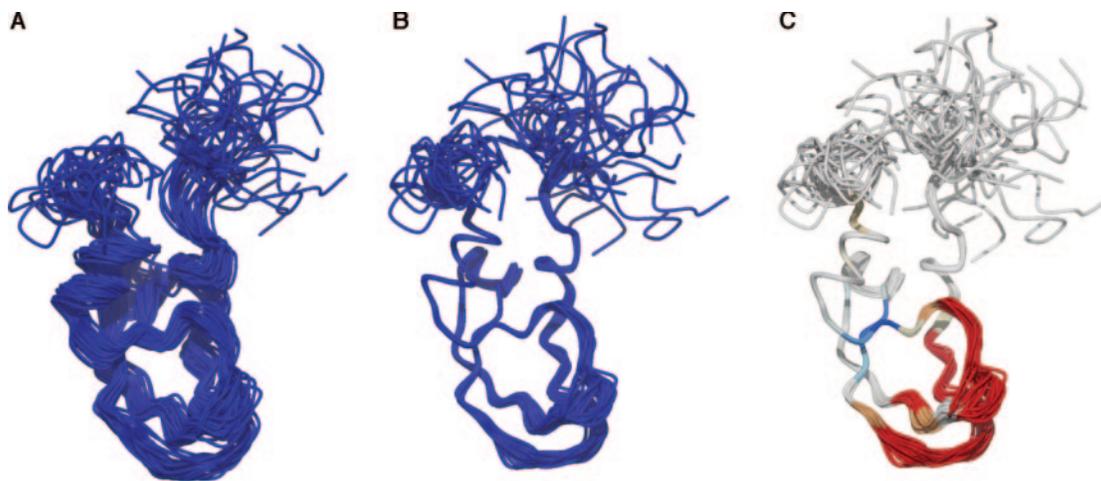
this assumption does not hold, a condition known in statistics as heteroscedasticity, LS can provide misleading and inaccurate results. However, the requirement for homogeneous variances is generally violated with macromolecular superpositions. For example, in reported superpositions of multiple NMR protein models the backbone variances commonly range over three orders of magnitude. Similarly, in comparisons of different protein domains belonging to the same fold, the structures deviate from each other with varying degrees of local precision: some atoms ‘superimpose well’ and others do not. LS further requires that the variances be uncorrelated. However, this assumption is also violated in the case of macromolecular superpositions. The variance for each atom is highly correlated with the variances of proximal atoms, owing to linkage resulting from inter-atomic chemical bonds and physical interactions.

To correct for these shortcomings of LS, we have applied the principle of maximum likelihood (ML) to the superposition problem by assuming a Gaussian distribution of the structures in the analysis (Theobald and Wuttke, 2006). ML is widely considered to be fundamental in statistical modeling and parameter estimation (Pawitan, 2001). ML superpositioning requires solving for four types of unknowns: a global covariance matrix describing the variance and correlations for each atom in the structures, a mean structure, and, for each structure in the analysis, a rotation matrix and a translation vector. In the present case, the ML method accounts for uneven variances and correlations in the structures by weighting by the inverse of the atomic covariance matrix. The unknowns are interdependent and cannot be solved analytically. For simultaneous estimation, we use an iterative numerical algorithm for maximizing the joint likelihood (see Supplementary data).

### 2 IMPLEMENTATION

Our numerical algorithm for calculating ML superpositions is implemented in the command-line UNIX program THESEUS. Rendered output is shown in Figure 1, where a comparison with the LS method clearly shows the increased accuracy of ML superpositions when including all atoms in the calculation. THESEUS works in two modes: (1) a mode for superpositioning structures with identical atoms and (2) an ‘alignment mode’ which can superposition homologous structures with different residues. Note that THESEUS is not a tool for structure-based sequence alignment, which is a separate bioinformatic challenge (Bourne and Shindyalov, 2003). Thus, like all structural superposition methods, THESEUS requires an a priori one-to-one mapping among the atoms/residues in the structures under consideration. When superpositioning multiple

\*To whom correspondence should be addressed.



**Fig. 1.** A conventional LS superposition versus the ML superposition (A and C) of 30 NMR models of the 71 amino acid Kunitz domain 2 of Tissue Factor Pathway Inhibitor (PDB ID: 1adz). All  $C_{\alpha}$ s were included in the calculations. For the LS superposition,  $\text{RMSD}_{\text{LS}} = 4.37$ , overall reduced  $\chi^2 = 27.9$ , absolute log likelihood =  $-9067.0$  and  $\text{AIC} = -9139.9$ . For the ML superposition,  $\text{RMSD}_{\text{ML}} = 0.113$ , the overall reduced  $\chi^2 = 1.01$ , absolute log likelihood =  $-1459.3$  and  $\text{AIC} = -1906.5$ . Relative to the ML superposition,  $\Delta\text{AIC} = 7177.8$ , indicating that the ML model is preferred by a large margin as judged by likelihoodist model selection criteria ( $P \approx 0.0$ , Vuong likelihood ratio test) (Burnham and Anderson, 1998; Vuong, 1989). Qualitatively similar results are seen with pairwise superpositions. The first principal component of the ML correlation matrix plotted on the Kunitz ML family superposition. The red-colored loops at lower right indicate regions that are strongly correlated within the family, whereas the light blue  $\beta$ -strands at middle left are modestly anti-correlated with the red regions.

conformations of the same protein (e.g. NMR models or different crystal structures of identical proteins), the one-to-one mapping is trivial. However, when superpositioning different proteins, the user must supply a sequence alignment of the proteins for THESEUS to use as a guide. THESEUS accepts sequence alignments in both CLUSTAL and A2M (FASTA) formats.

There is no limit on the number of structures that THESEUS will superposition (aside from that mandated by the operating system and memory capability). Via simple command line options, users can choose to superposition with the conventional LS method, to select residues (or alignment columns) for inclusion or exclusion from the calculation, and, when superpositioning structures of identical residues (mode 1), to select atom types (e.g. only  $\alpha$ -carbons or only backbone atoms). THESEUS writes out two PDB format files, one of the final superposition and one of the estimate of the mean structure. For easy visualization, the estimated variance for each atom is converted to a 'pseudo-B-factor' and written in the temperature factor field of the mean structure file.

In addition to estimating the optimal superposition of multiple structures, THESEUS calculates various frequentist and likelihood-based statistics for evaluating the fit and quality of the superposition, including the conventional least-squares  $\text{RMSD}_{\text{LS}}$ , the maximum likelihood  $\text{RMSD}_{\text{ML}}$ , and the reduced  $\chi^2$  for the overall superposition. The overall absolute likelihood is produced, as well as likelihoodist model selection measures such as the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) (Burnham and Anderson, 1998).

Finally, THESEUS will calculate the principal components of the covariance and correlation matrices for analysis of the major modes of correlated conformational differences within a superposition. Each principal component is written into the temperature factor field of two additional files: a superposition of all structures and the estimate of the mean structure. Principal components can then be visualized readily using software that

colors the structures according to values in the temperature factor field (Fig. 1C).

When assuming a diagonal covariance matrix (i.e. assuming no correlations), the calculation usually converges in a fraction of a second on modern personal computers for moderate-sized problems (e.g. 50 structures, 100  $\alpha$ -carbons) and in a few seconds for larger problems (e.g. 100 structures, 500  $\alpha$ -carbons). Calculation of the full atomic covariance matrix can take up to a few minutes for larger problems, as each iteration requires a matrix inversion.

## ACKNOWLEDGEMENTS

The authors are grateful to Olve Peersen for extensive bug-testing of THESEUS. The authors thank the NIH for funding (GM59414). D.L.T. is supported by Postdoctoral Fellowship Grant #PF-04-118-01-GMC from the American Cancer Society.

*Conflict of Interest:* none declared.

## REFERENCES

- Bourne, P.E. and Shindyalov, I.N. (2003) Structure comparison and alignment. In Bourne, P.E. and Weissig, H. (eds), *Structural Bioinformatics, Methods of Biochemical Analysis*. Wiley-Liss, Hoboken, NJ, Vol. 44, pp. 321–337.
- Burnham, K.P. and Anderson, D.R. (1998) *Model Selection and Inference: A Practical Information-Theoretic Approach*. Springer, New York.
- Flower, D.R. (1999) Rotational superposition: a review of methods. *J. Mol. Graph Model*, 17, 238–244.
- Pawitan, Y. (2001) *In All Likelihood: Statistical Modeling and Inference Using Likelihood*. Oxford Science Publications, Clarendon Press, Oxford.
- Seber, G.A.F. and Wild, C.J. (1989) Nonlinear regression. In *Wiley Series in Probability and Mathematical Statistics. Probability and Mathematical Statistics*. Wiley, New York.
- Theobald, D.L. and Wuttke, D.S. (2006) Empirical Bayes hierarchical models for regularizing maximum likelihood estimation in the matrix Gaussian Procrustes problem. *Proc. Natl Acad. Sci USA*, (In press).
- Vuong, Q.H. (1989) Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica*, 57, 307–333.

## SUPPLEMENTARY MATERIALS

### THEORY

#### A Gaussian statistical model for the superposition problem

A more general treatment of the following theoretical likelihood analysis can be found elsewhere in detail (Theobald and Wuttke, 2006). Here we briefly recap a simplification of the main results that specifically apply to macromolecular structural superpositioning as implemented in THESEUS.

Consider the case of superpositioning  $N$  different structures ( $\mathbf{X}_i$ ,  $i = 1 \dots N$ ), each with  $K$  corresponding atoms. We represent each structure as a  $K \times 3$  matrix of  $K$  rows of atoms, where each atom is a 3-vector.

We assume a perturbation model in which each macromolecular structure  $\mathbf{X}_i$  is distributed according to a matrix Gaussian probability density (Goodall, 1991; Lele, 1993). In this likelihood model, each observed structure  $\mathbf{X}_i$  is considered to be a randomly rotated and translated Gaussian perturbation of a mean structure  $\mathbf{M}$ :

$$\mathbf{X}_i = (\mathbf{M} + \mathbf{E}_i)\mathbf{R}'_i - \mathbf{1}_K \mathbf{t}_i \quad (1)$$

where  $\mathbf{t}_i$  is a  $1 \times 3$  translational row vector,  $\mathbf{1}_K$  denotes the  $K \times 1$  column vector of ones, and  $\mathbf{R}_i$  is a proper, orthogonal  $3 \times 3$  rotation matrix. The  $K \times 3$  matrix  $\mathbf{E}_i$  is a matrix of Gaussian random errors with mean zero, *i.e.*,  $\mathbf{E}_i \sim N_{K,3}(\mathbf{0}, \Sigma, \mathbf{I}_3)$ . Here  $\Sigma$  is a  $K \times K$  covariance matrix for the atoms, which describes the variance of each atom and the covariances among the atoms. For simplicity we assume that the variance about each atom is spherical.

#### The superposition likelihood equation

In general, the covariance matrix  $\Sigma$  is inestimable unless it is parametrically constrained. Therefore, to estimate the atomic covariance matrix we assume that its eigenvalues are distributed according to an inverse gamma distribution, which is physically reasonable for macromolecular structures. The joint log-likelihood for our likelihood superposition problem is thus the sum of the log-likelihood for the atomic covariance matrix eigenvalues and the log-likelihood for a multivariate matrix normal distribution (Arnold, 1981; Dutilleul, 1999) corresponding to the perturbation model described by Eq. 1. The full superposition log-likelihood  $\ell(\mathbf{R}, \mathbf{t}, \mathbf{M}, \Sigma; \mathbf{X}) = \ell_S$  is given by

$$\begin{aligned} \ell_S = & -\frac{1}{2} \sum_i^N \|(\mathbf{X}_i + \mathbf{1}_K \mathbf{t}_i)\mathbf{R}_i - \mathbf{M}\|_{\Sigma^{-1}}^2 \\ & -\frac{3NK}{2} \ln(2\pi) - \frac{3N}{2} \ln|\Sigma| \\ & -(1+\gamma) \ln|\Sigma| - \alpha \operatorname{tr} \Sigma^{-1} \\ & +K\gamma \ln \alpha - K \ln \Gamma(\gamma) \end{aligned} \quad (2)$$

where  $|\mathbf{U}|$  denotes the determinant of the matrix  $\mathbf{U}$ ,  $\|\mathbf{U}\|_{\mathbf{V}}^2 = \operatorname{tr}\{\mathbf{U}'\mathbf{V}\mathbf{U}\}$  denotes a squared Frobenius Mahalanobis matrix norm, and  $\alpha$  and  $\gamma$  are the scale and shape parameters, respectively, of an inverse gamma distribution of the atomic covariance matrix's nonzero eigenvalues ( $\lambda_j$ ):

$$P(\lambda_j) = \frac{\alpha^\gamma}{\Gamma(\gamma)} \lambda_j^{-(1+\gamma)} e^{-\frac{\alpha}{\lambda_j}} \quad (3)$$

#### ML superposition solutions

We provide in the following the ML solutions for each of the unknown parameters of the above superposition log-likelihood equation.

Each uncentered structure  $\mathbf{X}_i$  must be centered by translating it so that its row-weighted center is at the origin. Row-wise weighted centering is applied by

$$\tilde{\mathbf{X}}_i = \mathbf{X}_i + \mathbf{1}_K \hat{\mathbf{t}}_i \quad (4)$$

where  $\hat{\mathbf{t}}_i$  is the ML estimate of the optimal translation:

$$\hat{\mathbf{t}}_i = -\frac{\mathbf{1}'_K \Sigma^{-1} \mathbf{X}_i}{\mathbf{1}'_K \Sigma^{-1} \mathbf{1}_K}$$

The extended ML estimator of the inverse gamma distributed atomic covariance matrix  $\hat{\Sigma}_{I\gamma}$  is given by:

$$\hat{\Sigma}_{I\gamma} = \frac{3N}{3N + 2(\gamma + 1)} \left( \frac{2\alpha}{3N} \mathbf{I} + \hat{\Sigma}_U \right) \quad (5)$$

where the unrestricted ML estimator of the covariance matrix  $\hat{\Sigma}_U$  is:

$$\hat{\Sigma}_U = \frac{1}{3N} \sum_i^N (\tilde{\mathbf{X}}_i \mathbf{R}_i - \hat{\mathbf{M}})(\tilde{\mathbf{X}}_i \mathbf{R}_i - \hat{\mathbf{M}})' \quad (6)$$

Similarly, the extended ML estimator of the inverse gamma distributed eigenvalues  $\hat{\Lambda}_{I\gamma}$  is given by:

$$\hat{\Lambda}_{I\gamma} = \frac{3N}{3N + 2(\gamma + 1)} \left( \frac{2\alpha}{3N} \mathbf{I} + \hat{\Lambda}_U \right) \quad (7)$$

where  $\hat{\Lambda}_U$  is the diagonal matrix of eigenvalues of the unrestricted sample covariance matrix  $\hat{\Sigma}_U$ . The rotations are calculated via a singular value decomposition (SVD). Let the SVD of an arbitrary matrix  $\mathbf{D}$  be  $\mathbf{U}\mathbf{\Lambda}\mathbf{V}'$ . The optimal rotations  $\hat{\mathbf{R}}_i$  are then estimated by

$$\begin{aligned} \hat{\mathbf{M}}' \hat{\Sigma}^{-1} \tilde{\mathbf{X}}_i &= \mathbf{U}\mathbf{\Lambda}\mathbf{V}' \\ \hat{\mathbf{R}}_i &= \mathbf{V}\mathbf{P}\mathbf{U}' \end{aligned} \quad (8)$$

Rotoinversions are prevented by ensuring that the rotation matrix  $\hat{\mathbf{R}}_i$  has a positive determinant:  $\mathbf{P} = \mathbf{I}$  if  $|\mathbf{V}||\mathbf{U}| = 1$  or  $\mathbf{P} = \operatorname{diag}(1, 1, -1)$  if  $|\mathbf{V}||\mathbf{U}| = -1$ . The mean form is estimated as the arithmetic average:

$$\hat{\mathbf{M}} = \bar{\mathbf{X}} = \frac{1}{N} \sum_i^N \tilde{\mathbf{X}}_i \mathbf{R}_i \quad (9)$$

Finally, the overall "tightness" of a maximum likelihood superposition can be assessed by a maximum likelihood analog of the conventional least-squares root mean squared deviation ( $\operatorname{RMSD}_{LS}$ ):

$$\operatorname{RMSD}_{ML} = \sqrt{\frac{K}{\operatorname{tr} \hat{\Sigma}^{-1}}} \quad (10)$$

When all variances are equal and there are no correlations (*i.e.* the least-squares assumption), the two measures are equivalent.

## ALGORITHM

The ML solutions given above must be solved simultaneously using a numerical maximization algorithm. For this purpose, we have developed the following iterative algorithm based on the Expectation-Maximization (EM) method (Dempster *et al.*, 1977; Pawitan, 2001). In brief:

1. **Initialize:** Set  $\hat{\Sigma} = \mathbf{I}$ . Estimate the mean structure  $\hat{\mathbf{M}}$  using the EDMA embedding method (Lele, 1993).
2. **Translate:** Center each  $\mathbf{X}_i$  according to Eq. 4.
3. **Rotate:** Calculate each  $\hat{\mathbf{R}}_i$  according to Eq. 8, and rotate each centered structure accordingly:  $\mathbf{X}_i = \hat{\mathbf{X}}_i \hat{\mathbf{R}}_i$ .
4. **Estimate the mean:** Recalculate the average structure  $\hat{\mathbf{M}}$  according to Eq. 9. Return to step 3 and loop to convergence.
5. **Estimate the inverse gamma distributed eigenvalues  $\hat{\Lambda}$ :** Estimate  $\hat{\Sigma}_U$  from Eq. 6, and spectrally decompose it to find the sample eigenvalues  $\hat{\Lambda}_U$ . Estimation of the inverse gamma distributed eigenvalues  $\hat{\Lambda}_{I\gamma}$  involves the simultaneous solution of two problems: (1) the modification of  $\hat{\Lambda}_U$  according to Eq. 7 and (2) ML estimation of the scale and shape parameters of the inverse gamma distributed eigenvalues. Details of this step are given below in *Estimation of the eigenvalues*.
6. **Estimate the atomic covariance matrix  $\hat{\Sigma}$ :** Modify  $\hat{\Sigma}_U$  according to Eq. 5.
7. **Loop:** Return to step 2 and loop until convergence.

Note that if one assumes that the variances are all equal (*i.e.*, that  $\Sigma \propto \mathbf{I}$ ), then the above algorithm is equivalent to the conventional least-squares algorithm for multiple simultaneous superpositioning (Diamond, 1992; Gerber and Müller, 1987; Kearsley, 1990; Shapiro *et al.*, 1992).

### Estimation of the eigenvalues:

Estimation of the the diagonal matrix of inverse gamma distributed eigenvalues  $\hat{\Lambda}$  involves three steps: (1) calculate the unconstrained sample eigenvalue matrix  $\hat{\Lambda}_U$  by eigen decomposition of the sample covariance matrix given by Eq. 6, (2) estimate the scale and shape parameters of an inverse gamma distribution based on these eigenvalues, and (3) modify the unconstrained sample eigenvalues  $\hat{\Lambda}_U$  according to Eq. 7. Steps 2 and 3 must be performed simultaneously. An iterative EM algorithm that performs this simultaneous estimation is described below.

Before iterating, the EM algorithm must be initialized. In general, the sample covariance matrix (Eq. 6) is first spectrally decomposed to determine the sample eigenvalues. However, this decomposition is unnecessary if one assumes that the covariance matrix is diagonal, since then the variances are the eigenvalues (with an important exception described below). In either case, the scale and shape parameters of the inverse gamma distribution are estimated based on the non-zero, positive eigenvalues. Note that the sample covariance matrix is always rank degenerate, *i.e.*, there are always multiple zero eigenvalues, regardless of the number of structures used in the calculation. Because of the nature of the superposition problem, and usually due to insufficient data, the sample covariance matrix is of maximum rank  $K - 3$ , and may even be less when there are few structures ( $\text{rank} = \min(3N - 6, K - 3)$ ). We treat these missing eigenvalues as missing data, and in the EM algorithm the inverse gamma fit is based only on the  $\min(3N - 6, K - 3)$  full rank, positive eigenvalues. When assuming that the covariance matrix is

diagonal (no correlations), it is then necessary to omit the smallest three variances from the inverse gamma fit, as they are known *a priori* to be a zero-valued eigenvalues.

The algorithm cycles until convergence between the following two steps:

1. **Fit the inverse gamma parameters:** Find the ML estimates of the inverse gamma scale and shape parameters for the current eigenvalues. A maximum likelihood fit to an inverse gamma distribution can be accomplished by taking the inverse of each data point and fitting the transformed data to a gamma distribution (Evans *et al.*, 2000). For example, if  $\mu_i = \lambda_i^{-1}$  for all positive eigenvalues, then the corresponding gamma probability distribution is:

$$P(\mu_j) = \frac{\alpha^\gamma}{\Gamma(\gamma)} \mu_j^{(\gamma-1)} e^{-\alpha\mu_j}$$

The ML estimates of the parameters  $\alpha$  and  $\gamma$  are then solutions of the simultaneous equations:

$$\hat{\alpha} = \frac{\gamma}{\bar{\mu}}$$

$$y = \ln \hat{\gamma} - \psi_{(0)}(\hat{\gamma}) - \ln \bar{\mu} + \frac{1}{K'} \sum_j^{K'} \ln \mu_j = 0$$

where  $K'$  is the number of nonzero eigenvalues,  $\bar{\mu} = \sum_j^{K'} \mu_j / K'$ , and  $\psi_{(0)}$  is the digamma function. Newton's method can be used readily to solve the last equation above using its first derivative:

$$\frac{\partial y}{\partial \hat{\gamma}} = \frac{1}{\hat{\gamma}} - \psi_{(1)}(\hat{\gamma})$$

where  $\psi_{(1)}$  is the trigamma function (the first derivative of the digamma function). For the first iteration, useful starting values for inverse gamma parameters in the Newton method fit are given by the method of moments estimators,  $\hat{\alpha} = \bar{\mu} / \phi$  and  $\hat{\gamma} = \bar{\mu}^2 / \phi$ , where  $\phi$  is the variance of the nonzero eigenvalues. In subsequent iterations, starting values in the Newton method are given by the parameter values from the previous iteration.

2. **Modify the sample eigenvalues:** Modify the sample eigenvalues according to Eq. 7.

## REFERENCES

- Arnold, S. F. (1981) *The Theory of Linear Models and Multivariate Analysis*. Wiley, New York.
- Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977) Maximum likelihood from incomplete data via the EM algorithm. *J Roy Stat Soc B Met*, **39**, 1–38.
- Diamond, R. (1992) On the multiple simultaneous superposition of molecular-structures by rigid body transformations. *Protein Sci*, **1**, 1279–1287.
- Dutilleul, P. (1999) The MLE algorithm for the matrix normal distribution. *J Stat Comput Sim*, **64**, 105–123.
- Evans, M., Hastings, N. and Peacock, J. B. (2000) *Statistical Distributions*. Wiley series in probability and statistics. Probability and statistics. John Wiley and Sons, New York, 3rd edition.
- Gerber, P. R. and Müller, K. (1987) Superimposing several sets of atomic coordinates. *Acta Crystallogr A*, **43**, 426–428.
- Goodall, C. (1991) Procrustes methods in the statistical analysis of shape. *J Roy Stat Soc B Met*, **53**, 285–321.
- Kearsley, S. K. (1990) An algorithm for the simultaneous superposition of a structural series. *J Comput Chem*, **11**, 1187–1192.

---

Lele, S. (1993) Euclidean distance matrix analysis (EDMA) - estimation of mean form and mean form difference. *Math Geol*, **25**, 573–602.

Pawitan, Y. (2001) *In All Likelihood: Statistical Modeling and Inference Using Likelihood*. Oxford Science Publications. Clarendon Press, Oxford.

Shapiro, A., Botha, J. D., Pastore, A. and Lesk, A. M. (1992) A method for multiple superposition of structures. *Acta Crystallogr A*, **48**, 11–14.

Theobald, D. L. and Wuttke, D. S. (2006) Empirical Bayes hierarchical models for regularizing maximum likelihood estimation in the matrix Gaussian Procrustes problem. *PNAS*, **in press**.