An Algorithm for the Simultaneous Superposition of a Structural Series

Simon K. Kearsley

Merck Sharp & Dohme Research Laboratories, Rahway, New Jersey 07065

A procedure is delineated for finding the optimal superposition of a series of chemical structures. Quaternions are used to represent the absolute orientation of the structures—rotations of the structures need never be carried out explicitly. The Rational Function Optimization method is used to minimize a simultaneous superposition residual similar to the one given by Gerber and Müller.[1] The robustness of the method is illustrated by comparing a series of conformations of a polyene carotenoid.

INTRODUCTION

The task of making comparisons between a series of related structures is ubiquitous among chemists, crystallographers, biochemists, and pharmacologists. These scientists are constantly building models and aligning them to garner some measure of structural or chemical similarity. The events that lead to the comparison between structures can be broken up into two phases. The first phase is the most difficult; superposition requirements are often subjective and depend on the type of comparison one wants to attain. For instance, what is more important, steric overlap of atoms or the overlap of like partial charges? For most atom based superposition schemes, one must have readily available atom-atom associations between the structures. The methods by which these are obtained will not be broached in this article. The second phase is procedural and actually carries out the alignment, either by reporting back the value of the alignment function or presenting the required transformations that will eventually be involved with a graphical visualization. The goal of this article is to present a technique of simultaneously superposing a series of structures where the atom-atom associations have been previously defined between all pairs of structures.

   Comparing a series of related molecules or a collection of structural conformations can be done in a pairwise fashion. One structure is chosen to remain static and each of the rest of the structures are individually superposed to minimize the sum of the distances squared between corresponding atoms of these two structures. However, optimal superposition between pairs of structures not involved with the static structure is not guaranteed. To circumvent this Gerber and

Müller[1] minimize a simultaneous superposition residual of the form similar to eq. (1).

$$RS = \sum_{i}^{n-1} \sum_{j=i+1}^{n} \sum_{k}^{ns} (Q^i \mathbf{x}_k^i - Q^j \mathbf{x}_k^j)^2 \qquad (1)$$

Here, the superposition of all possible pairs of structures is optimized simultaneously. The first two sums are over the unique pairs of structures, the third is over the number of atom-atom associations between the structures. The root mean square deviation in distance is therefore given by RMS = $\sqrt{2RS/(ns.n(n-1))}$. The chief differences between the method presented here and that of Gerber and Müller is the technique of optimization and the parameterization of the rotation operator Q. Also, during the calculation the atomic coordinates of the structures are never altered. This means that atoms not involved in the superposition need not be carried along with each structure—the algorithm remains uncluttered by uninvolved data. After the computations are complete the transformation operators can be applied to each entire structure.

   Applying this superposition coordinate transformation allows for the calculation of an average structure. Gerber and Müller[1] give a prescription for this. Also, they have included a general weighting scheme in their simultaneous superposition residual. It would be a simple task to modify the computer code to accommodate their weighting scheme, but more often than not, such a weighting scheme is never used. In this paper we are assuming unit weights for all structures and atom-atom associations.

METHODS

In both the current method, and in that of Gerber and Müller, one structure is selected as the static

structure and the preliminary pairwise superposition onto this structure is accomplished by the established analytic methods.[2-4] This involves shifting all the centroids to the origin of coordinates and then performing an orthogonal transformation that minimizes the sum of the distances squared between associated atoms. This transformation usually orients the structures near the global minimum,[5] the lowest possible value for the simultaneous superposition residual

$$
Q = \begin{bmatrix} (q_1^2 + q_2^2 - q_3^2 - q_4^2)/p & 2(q_2q_3 + q_1q_4)/p & 2(q_2q_4 - q_1q_3)/p \\ 2(q_2q_3 - q_1q_4)/p & (q_1^2 + q_3^2 - q_2^2 - q_4^2)/p & 2(q_3q_4 + q_1q_2)/p \\ 2(q_2q_4 + q_1q_3)/p & 2(q_3q_4 - q_1q_2)/p & (q_1^2 + q_4^2 - q_2^2 - q_3^2)/p \end{bmatrix} \tag{2}
$$

function in eq. (1). For the simultaneous residual it is a straightforward task to show that optimal superposition with respect to translation of the structures requires the centroids remain superposed. Where there is digression from the method of Gerber and Müller is the technique of optimization. They chose to linearize their representation of the rotation operator and solve it like a linear least squares problem. Instead, this method represents the rotation operator as a function of quaternion parameters and optimizes the residue using the Rational Function Optimization (RFO) technique. This procedure is expected to be more stable for pathological applications, as illustrated by the rubixanthin example, where preliminary orientations are far from the global minimum. The next few sections introduce quaternions and the RFO method.

Quaternion Parameterization

In 1843, Sir William R. Hamilton invented the algebra of quaternions; these can be thought of as quadruplets of real numbers. He soon found out that quaternions could parameterize rotations. According to Altmann[6] though, in 1840, the nearly forgotten and largely uncredited Olinde Rodriques derived, in a thorough geometric exposition, an equivalent parameterization of rotations. Unfortunately, this parameterization later got attributed to Euler, sometimes called the Euler–Rodriques' parameters. The much older three Eulerian angles (ϕ, θ, φ), distinct from Euler–Rodriques' parameters or quaternions, have been the main staple of scientists for the parameterization of rotation—that is, describing the orientation of structures. Recently however, there has been a resurgence in the use of the quaternionic parameterization. There are two main reasons for this. Firstly, computers work faster with algebraic rather than trigonometric functions—this is aside from the obvious elegance of quaternion algebra over Euler angle ma-

nipulations. Secondly, derivative space is continuous for quaternions (at least as far as the first and second derivatives) whereas Euler angles show a plane of singularities when the azimuthal angle θ nears zero or π. Thus numerical optimizers that work with derivatives of quaternions will show superior convergence properties to Euler angle parameterizations.[7,8]

With the four quaternion parameters [q_1, q_2, q_3, q_4], the rotation operator Q is given by:

where $p = q_1^2 + q_2^2 + q_3^2 + q_4^2$ is the norm squared of the quaternion and explicitly normalizes Q. If the quaternions were not normalized, then the extra degree of freedom would have the effect of isotropically dilating or shrinking the structures. To prevent this the quaternions are restrained by a normalizing penalty function of the form:

$$
\text{Penalty} = \frac{1}{2}g(1 - p)^2 \tag{3}
$$

The constant g scales the penalty function. Although the norm of the quaternions are allowed to adjust away from unity, in practice, they are always normalized after each optimization iteration. Therefore, first derivatives of the normalizing penalty function are always zero and the effect of the penalty function is only to level shift the second derivative (Hessian) eigenvalues that deal with dilation (or shrinking) out of the way of the eigenmodes pertaining to rotation. This is reminiscent of the level shifting technique used to apply the Eckart constraints in molecular mechanics programs that utilize quadratic minimizers.[9] Other advantages of the quaternionic parameterization is that Q is a pure algebraic function, there are no trigonometric terms, so taking first and second derivatives of Q with respect the quaternion components is straightforward. Also, proper rotations are always produced; the determinant of Q is always positive regardless of the values taken by the quaternion components.

Rational Function Optimization

The Rational Function Optimization approach was used to optimize parameters of the electronic wave function for ground and excited states[10] and extended to facilitate surface walking techniques on potential energy surfaces.[11] First we shall introduce the RFO method and then describe why this is preferable to other quadratic optimizers in this application.

The difference residual ΔRS can be expanded in a Taylor series up to the quadratic term about

the current structure orientations as follows:

$$\Delta RS(\mathbf{q}) = RS(\mathbf{q} + \mathbf{q}_0) - RS(\mathbf{q}_0)$$

$$= \mathbf{g}^T\mathbf{q} + \frac{1}{2}\mathbf{q}^T H\mathbf{q} \qquad (4)$$

where $\mathbf{g}$ and H are the first and second derivatives respectively, of RS at $\mathbf{q}_0$. The shift vector $\mathbf{q}$ adjusts the quaternion components and therefore the orientation for all the structures. The original derivation of the RFO approach chose to temper the quadratic approximation of eq. (4) by subtracting a correcting function of the form $\Delta RS(\mathbf{q})\mathbf{q}^T S\mathbf{q}$. The effect of this is that as $\mathbf{q}$ becomes large ΔRS will tend to a limiting value rather than infinity.[10] Alternatively (and admittedly rather in hindsight), ΔRS could have been expanded as a rational function or quotient of polynomials; eq. (5) shows a [2, 2] form with quadratic polynomials in the numerator and denominator.

$$\Delta RS(\mathbf{q}) = \frac{p_0 + \mathbf{p}_1^T\mathbf{q} + \mathbf{q}^T p_2 \mathbf{q}}{r_0 + \mathbf{r}_1^T\mathbf{q} + \mathbf{q}^T r_2 \mathbf{q}} \qquad (5)$$

The [2, 2] Padé approximant is a special form of the rational function in which the power series expansion of eq. (5) will match the first few terms of the straight forward Taylor series expansion of the residual.[12] To do this, put eq. (4) in the numerator, set the scaler r_0 equal to 1 (r_0 is quite arbitrary), set $\mathbf{r}_1$ to 0 and replace r_2 with the symmetric square matrix S to arrive at eq. (6).

$$\Delta RS(\mathbf{q}) = \frac{\mathbf{g}^T\mathbf{q} + \frac{1}{2}\mathbf{q}^T H\mathbf{q}}{1 + \mathbf{q}^T S\mathbf{q}} = \frac{\frac{1}{2}[\mathbf{q}^T \; 1]\begin{bmatrix} H & \mathbf{g} \\ \mathbf{g}^T & 0 \end{bmatrix}\begin{bmatrix} \mathbf{q} \\ 1 \end{bmatrix}}{[\mathbf{q}^T \; 1]\begin{bmatrix} S & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} \mathbf{q} \\ 1 \end{bmatrix}} \qquad (6)$$

The expansion of $(1 + \mathbf{q}^T S\mathbf{q})^{-1}$ only contributes to the third and higher order terms and so the first and second derivatives of ΔRS remain the same as those of the Taylor series. The matrix S was originally used to describe the mixing of wave functions or overlap.[10] It could in principle be used to improve the conditioning of the Hessian or at least rescale the variables so that the residual surface is less elliptical. That is, unit movement in different variables will produce changes of similar absolute magnitude in the residual function. In most applications S is best kept to the unit matrix and for the purposes of this article S will be the unit matrix.

The trick to the RFO procedure is to rewrite the Padé approximant in matrix form as shown in eq. (6). Multiplying through by the denominator, differentiating and setting to zero to apply the stationary condition $\partial \Delta RS/\partial \mathbf{q} = 0$ to eq. (6), yields an eigenvalue equation of dimension one

more than the Hessian.

$$\begin{bmatrix} H & \mathbf{g} \\ \mathbf{g}^T & 0 \end{bmatrix}\begin{bmatrix} \mathbf{q} \\ 1 \end{bmatrix} = 2\Delta RS\begin{bmatrix} S & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} \mathbf{q} \\ 1 \end{bmatrix} \qquad (7)$$

Let the matrix combination of the H and $\mathbf{g}$ be called the RFO matrix. The shift vector $\mathbf{q}$ that adjusts the residual function towards a minimum is given by the intermediately normalized eigenvector (the eigenvector divided through by the last element) corresponding to the smallest eigenvalue of the RFO matrix, that is, the largest most negative eigenvalue. The eigenvalue gives twice the quadratic change in the residual function. Following along other eigenvectors will not generate the greatest decrease in the residual function and will probably increase it. Surface walking techniques take advantage of this behavior by moving away from minima and veering towards stationary points characterized by one or more negative eigenvalues.[11]

The RFO method is a *global* minimizer.[5] It strives towards stationary points on the residual surface that have all their Hessian eigenvalues positive–minima. It will not converge onto stationary points with maxima and is therefore superior to the Newton method in this application. A global minimizer ensures that every component of the shift vector, when taken in isolation, will reduce the value of the residual function and the next paragraph will show implicitly that this is true for the RFO method. Near the minimum, the optimal length of the RFO shift vector is given by the intermediately normalized eigenvector. This has been shown to produce quadratic convergence.[10]

In comparison with other optimization techniques, the RFO method is most similar to the well established step-length constrained Newton[13] or the Levenberg–Marquardt nonlinear least squares methods.[14] These techniques ensure *global* convergence and seek to increase the convergence radius of the Newton procedure. They do this by level shifting the negative eigenmodes that produce movement in the wrong direction (negative eigenvalues produce shift directions towards maxima) by making these recalcitrant eigenvalues positive. Whether this is done by the Lagrange multiplier technique that applies a step-length constraint $\partial\frac{1}{2}\lambda(r^2 - \mathbf{q}^T\mathbf{q})/\partial\mathbf{q} = 0$, where r^2 is the square of the required step-length, or more direct perturbation of the Hessian to give the shift vector more steepest descent character, the general result is that a multiple of the unit matrix $-\lambda S$ is added to the Hessian H:

$$(H - \lambda S)\mathbf{q} = -\mathbf{g} \qquad (8)$$

With a little rearrangement, the stationary condition for eq. (6) can be cast into the form of eq. (8)

with $\lambda = 2\Delta RS$. Thus the RFO method is a special case of modified-Hessian methods with the level shifting parameter λ having twice the change in quadratic residual function.

Choosing λ is critical, if it is too small then some negative eigenmodes may not change sign and the shift direction will become erroneous. Alternatively, if λ is too large convergence properties will suffer. The advantage the RFO method has over the other methods is that the level shifting parameter is determined implicitly by the algebra of the method. Furthermore, only the smallest eigenvalue and eigenvector need be determined to attain the shift vector $[\mathbf{q}, 1]$. For large problems, partial diagonalization methods such as coordinate relaxation[15] or the method of optimal relaxations[16] can be employed where the required effort becomes proportional to the square of the size of the RFO matrix, thus making it competitive to conjugate gradient or variable metric methods for timings per iteration. In general, if the calculation of the second derivative matrix is not particularly costly, then the RFO method will be faster than the conjugate and variable metric methods due to the accuracy and larger magnitude of its shift vector.

Returning to this particular application, the residual function RS has finite values for all possible values of Q^i so the expectation is that the RFO approximant, which also remains finite for all $\mathbf{q}$, will do a better job of following the RS surface. This being so, intuition would lead one to believe that the convergence radius will also be larger, and it is well documented that the range of convergence of the $[N, N]$ Padé approximants greatly exceeds the partial sums of Taylor series.[12]

The Line Search

Even for the RFO method, as with most optimizers, it is necessary to judge how far along the new shift vector to move. This suboptimization problem is called the line search. There are many ways of conducting a line search and the choice depends on how much time and effort the problem merits to arrive at the optimum magnitude of $\mathbf{q}$. The RFO method signals the approach to a minimum when the absolute value of the last element of the eigen/shift vector approaches unity (when it is unity, $\mathbf{q}$ is zero and convergence is achieved). In this application a rudimentary line search is achieved by scaling the eigenvector by division through by the last element providing the absolute value of the last element is above a threshold value of 0.9. Otherwise scaling is accomplished by performing binary chops on the shift vector magnitude until changes in orientation for the structures are less than 10 degrees. The change in orientation, calculated from the quaternion that

relates the current to the previous orientation, is given by eq. (9) where the superscipts C and P on the normalized quaternion components refer to the current and previous orientations respectively.[6,7]

$$\Delta\theta = 2 \cos^{-1}(q_1^C q_1^P + q_2^C q_2^P + q_3^C q_3^P + q_4^C q_4^P)$$
$$(9)$$

In addition to the maximal 10 degree movement of the structures, the current residual value is checked against its previous value. If the previous iteration had overstepped and the residual value increased, an unlikely occurrence for routine applications, then further binary chops are conducted on subsequent steps until the residual value drops below the previous value thus ensuring minimization. This check on the residual value is only needed on rare occasions.

Implementation in FORTRAN

The optimization algorithm for eq. (1) is contained in the subroutine COMSUP and has been sent to QCPE.[17] The preliminary pairwise orientations, and hence starting quaternions, are calculated using Kearsley's method.[4] It is expected that this subroutine will be subsumed into an existing modeling package that will provide the coordinates of the matching atoms. The output will consist of a set of rotation matrices and centering vectors that are to be applied to each structure.

EXAMPLE

The conformations of a natural product carotenoid, extracted from rosehip berries known as rubixanthin,[18] are used to illustrate the robustness of the method. The Cartesian coordinates of the polyene were constructed using CONCORD.[19] The conformations were generated using standard distance geometry techniques.[20,21] All double bonds were held in a planar trans configuration. The two ends of the molecule (the hydroxyl oxygen and the starred carbon atom) were held at a distance of 30 Å; this is approximately 2 Å shorter than the fully extended conformation generated by CONCORD. A slack of 0.3 Å was allowed for conformations to be accepted.

Fifty conformations for rubixanthin (A001 to A050) were generated and then compared. The value of g, the quaternion level shifting parameter, was set at 10,000. Convergence was achieved for the simultaneous residual when the norm of the gradient fell below 10^{-6}. Several of the structures were chosen to remain static and the initial pairwise superposition was used as a starting point for the optimization of the simultaneous residual. Table I shows the results for the 50 conformers. Figure 1 shows a similar but less dra-

Table I. Results of rubixanthin superpositions. The values of the pair superpositions are given as the overall root mean square deviation in Ångstroms. The final simultaneous residual minimum had a value of 1.7473 Ångstroms.

Static structure	Pair superposition	Average[a]	Maximum[a]	Cycles[b]
A001	1.7828	17.8	55.2	7
A010	1.8261	27.7	99.0	21
A020	1.8381	46.6	179.3	51
A030	1.8409	35.6	120.1	16
A040	1.8336	33.0	172.9	34
A050	1.8178	36.9	174.5	30

[a]The average and maximum angular difference between the pairwise and simultaneous superposition methods as calculated by eq. (9) and measured in degrees.
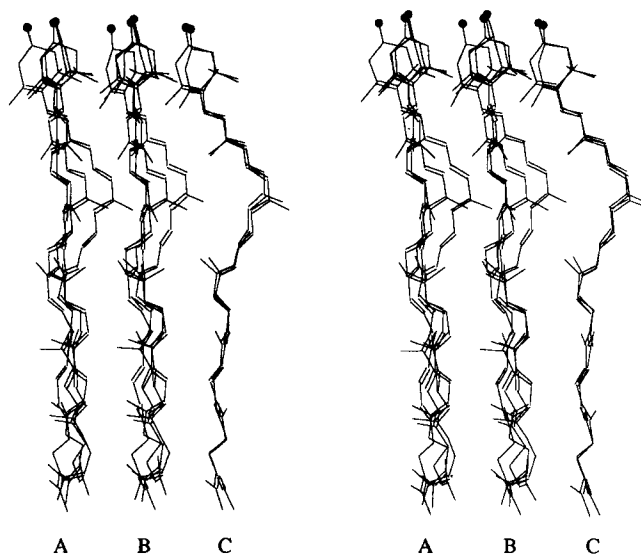[b]The number of iteration cycles needed to achieve convergence.

Figure 1. Stereo diagram showing the superposition of four conformers of rubixanthin. **A** shows the pairwise superposition (residual of 1.6427 Å), **B** shows the simultaneous superposition (residual of 1.6016 Å) and **C** compares the conformer with the maximum angular difference between the two methods (23.4 degrees).

matic comparison for only four of the conformations, A001 to A004.

For the 50 conformer comparison, all the pairwise starting points converged to the same minimum; this has a residual value of 1.7473 Å. The average change from the pairwise residual to the simultaneous residual minimum is 0.076 Å which might not seem very remarkable until the number of matching interactions is considered. There are 41 nonhydrogen atoms in rubixanthin and 50 conformers; this means there are 41 × 50 × (50−1)/2 = 50225 distances minimized, and therefore a total drop of approximately 3,800 Å for the sum of these distances. More significantly, the average angular changes calculated between the pairwise and simultaneous superposition

methods are very substantial. These angle differences were calculated using eq. (9). Indeed, in some cases the conformer with the greatest change in orientation had undergone a complete about face. For instance, choosing A020 to remain static resulted in a particularly bad pairwise starting orientation which required one of the structures to rotate by 179 degrees. In spite of this, the same (and probably global) simultaneous superposition residual minimum was achieved, thus demonstrating the robustness of the optimization algorithm and versatility of the quaternion parameterization of rotation.

Rubixanthin was chosen to illustrate how the algorithm fared with rather bad starting orientations. The reason that the conformers of rubixan-

Rubixanthin
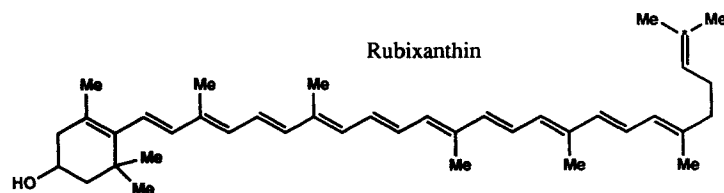
thin are difficult to align is due to their elongated shapes which were constrained to be approximately linear. The position of the pairwise minima along two of the rotational degrees of freedom, those along the short axes of the static molecule (axes corresponding to the two larger moments of inertia) are obviously congruent for the superposed molecules. However, the position of the pairwise superposition minima about the long axis of the static molecule bear little relationship between aligning molecules and can be as far from each other as 180 degrees. For more globular shaped molecules a syllogistic reckoning is often sufficient for pairwise superpositions. That is, if structures A and B are optimally superposed on structure C, it usually means that A and B are therefore near optimum superposition themselves. Hence, to attain the simultaneous superposition orientation for structures A, B, and C requires only minor adjustments (less than 5 degrees); this can be achieved within two or three optimization cycles.

References

1. P. R. Gerber and K. Müller, *Acta Crystallographia,* **A43,** 426 (1987).
2. W. Kabsch, *Acta Crystallographia,* **A34,** 827 (1978).
3. R. Diamond, *Acta Crystallographia,* **A44,** 211 (1988).
4. S. K. Kearsley, *Acta Crystallographia,* **A45,** 208 (1989).
5. A *global minimum* means the lowest possible value of the residual function. A *global minimizer* or the property of *global convergence* means that the optimization method finds minima rather than generic stationary points—that is, stationary points characterized by a positive-definite Hessian.
6. S. L. Altmann, *Rotations, Quaternions and Double Groups,* Oxford University Press, New York, 1986.
7. D. J. Evans, *Molecular Physics,* **34,** 317 (1977).
8. A. O. Griewank, B. R. Markey, and E. J. Evans, *J. Chem. Phys.,* **71,** 3449 (1979).
9. D. T. Nguyen and D. A. Case, *J. Phys. Chem.,* **89,** 4020 (1985).
10. R. Shepard, I. Shavitt and J. Simons, *J. Chem. Phys.,* **76,** 543 (1982).
11. A. Banerjee, N. Adams, J. Simons and R. Shepard, *J. Chem. Phys.,* **89,** 52 (1985).
12. G. A. Baker and J. L. Gammel, *The Padé Approximant in Theoretical Physics,* Academic Press, New York, 1970.
13. R. Fletcher, *Practical Methods of Optimization,* Wiley, New York, 1980.
14. D. W. Marquardt, *J. Soc. Ind. Appl. Math.,* **11,** 431 (1963).
15. D. K. Faddeev and V. N. Faddeeva, *Computational Methods of Linear Algebra,* (English Translation), Section 61, W. H. Freeman & Co., San Francisco, CA, 1963.
16. I. Shavitt, C. F. Bender, A. Pipano and R. P. Hosteny, *J. Computational Phys.,* **11,** 90 (1973).
17. Quantum Chemistry Program Exchange, Department of Chemistry, Indiana University, Bloomington, Indiana 47405, U. S. A.
18. B. O. Brown and B. C. L. Weedon, *Chem. Commun.,* 382 (1968).
19. CONCORD is a program for the automatic creation of 3D molecular structures from connection table information. Developed by R. S. Pearlman. Distributed by TRIPOS Associates, Inc.
20. Program DGEOM written by J. D. Andose, Merck & Co., Inc., as part of an in-house molecular modeling system is based on subroutine EMBEDD, written by G. M. Crippen.
21. G. M. Crippen, *Distance Geometry and Conformational Calculations,* John Wiley & Sons, New York, N.Y., 1981.