Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Map self-validation: improved criteria to resolve the SIR or SAS phase ambiguity

David A. Langs,* Robert H. Blessing and Dongyao Guo

Hauptman–Woodward Medical Research Institute Inc., 73 High Street, Buffalo, NY 14203, USA

Correspondence e-mail: langs@algol.hwi.buffalo.edu

Received 25 April 2001
Accepted 11 July 2001

A procedure was recently described that used the correlation coefficient (CC) agreement between the observed $|F_{\mathbf{h}}|$ and their associated unbiased 'omit map' extrapolated values $|X_{\mathbf{h}}|$ from an initial trial map as the basis for resolving the SIR or SAS phase ambiguity. It is noted here that a significant improvement in selectivity can be obtained if this agreement is expressed in terms of the complex-valued $F_{\mathbf{h}}$ and $X_{\mathbf{h}}$. A new scheme is outlined to exploit the weighted average of the two SIR or SAS phase choices. This procedure requires six FFTs per phase compared with three for the older method that randomly selected either of the two permitted phase choices from the Argand diagram as starting values. Trial calculations are encouraging for applications as low as 4 Å resolution.

1. Background

A previous paper (Langs *et al.*, 2001) described a validation method whereby a set of unbiased structure-factor estimates, $X_{\mathbf{h}}$, could be efficiently computed from an initial trial map. Values for each $X_{\mathbf{h}}$ are in principle extrapolated from a series of density-modified maps, from each of which in turn the associated $F_{\mathbf{h}}$ term has been excluded.

This procedure thus requires two FFTs for each estimated $X_{\mathbf{h}}$: the first to compute the map for which any particular $F_{\mathbf{h}}$ has been excluded and the second to back-transform this map after it has been modified to obtain the value $X_{\mathbf{h}}$. Since macromolecular data sets generally include thousands of independent measured $F_{\mathbf{h}}$ data, such brute-force calculations to obtain the $X_{\mathbf{h}}$ values are clearly out of the question. Remarkably, in its place, a Fourier convolution relationship was used to obtain the full set of $X_{\mathbf{h}}$ values in only three FFTs: (i) $\rho(\mathbf{r})$ is computed for the full set of $F_{\mathbf{h}}$ values; $\rho(\mathbf{r})$ is modified to $\rho'(\mathbf{r})$ by zeroing the density below some low threshold value, say $0.25\sigma(\rho)$; (ii) $\rho'(\mathbf{r})$ is back-transformed to obtain biased estimates, $F'_{\mathbf{h}}$, of $F_{\mathbf{h}}$; $\mu(\mathbf{r})$ is a mask of $\rho'(\mathbf{r})$ that equals 1.0 if $\rho'(\mathbf{r}) \neq 0$; (iii) $\mu(\mathbf{r})$ is back-transformed to obtain its Fourier coefficients $G_{\mathbf{h}}$. The defining equation for each $X_{\mathbf{h}}$ is

$$X_{\mathbf{h}} = F'_{\mathbf{h}} - (1/V)\sum_{\mathbf{h}_j} F_{\mathbf{h}_j} G_{\mathbf{h}-\mathbf{h}_j}, \qquad (1)$$

where the summation over $\mathbf{h}_j$ back-corrects each $F'_{\mathbf{h}}$ for the symmetry-related forms of $F_{\mathbf{h}}$ that must be excluded from $\rho(\mathbf{r})$ prior to estimating each $X_{\mathbf{h}}$.

Note that the variance, $\sigma^2(\rho)$, associated with any electron-density map is in the limit of integration independent of the phases of the data. It need not be computed from the actual grid-point densities of the map, but rather the sum of the $|F_{\mathbf{h}}|^2$ themselves in accordance with Parseval's equality (Read, 1997),

$$\begin{aligned}\sigma^2(\rho) &= \textstyle\int_V \rho(\mathbf{r})^2\,\mathrm{d}V / \int_V \mathrm{d}V \\ &= \textstyle\sum_{\mathbf{h}} |F_{\mathbf{h}}|^2 / V^2. \end{aligned} \qquad (2)$$

It is possible to compare any two phase-related maps, that is two maps for which all the phases are identical except one, which is assigned a different value, φ_1 *versus* φ_2, in each map. The CC agreement between the full set of $F_{\mathbf{h}}$ and $X_{\mathbf{h}}$ values is often a good indicator as to which of the two phase values for a particular $F_{\mathbf{h}}$ is best. The better map usually has the larger CC value and the best phase indications are generally associated with the largest differences $|\delta\mathrm{CC}| = |\mathrm{CC}_{\varphi 1} - \mathrm{CC}_{\varphi 2}|$ between both CC values. In this regard, we previously used

$$\mathrm{CC}_R = \frac{(\langle|F_{\mathbf{h}}X_{\mathbf{h}}|\rangle - \langle|F_{\mathbf{h}}|\rangle\langle|X_{\mathbf{h}}|\rangle)}{[(\langle|F_{\mathbf{h}}|^2\rangle - \langle|F_{\mathbf{h}}|\rangle^2)(\langle|X_{\mathbf{h}}|^2\rangle - \langle|X_{\mathbf{h}}|\rangle^2)]^{1/2}}, \qquad (3)$$

which was based on the magnitudes of $F_{\mathbf{h}}$ and $X_{\mathbf{h}}$. If, however, we consider evaluating this CC agreement in real space rather than reciprocal space by computing the density

$$\rho''(\mathbf{r}) = \textstyle\sum X_{\mathbf{h}} \exp(-2\pi i \mathbf{h}\cdot\mathbf{r}), \qquad (4)$$

it can be shown (Read, 1986) that

$$\mathrm{CC}_\rho = \int_V \rho(\mathbf{r})\rho''(\mathbf{r})\,\mathrm{d}V/\sigma(\rho)\sigma(\rho'')\int_V \mathrm{d}V \quad (5)$$
$$= \textstyle\sum_{\mathbf{h}} |F_{\mathbf{h}}X^*_{\mathbf{h}}|\cos(\varphi_{\mathbf{h}} - \psi_{\mathbf{h}})/\sigma(\rho)\sigma(\rho'')V,$$

which involves taking the real value of the sum of the product of complex structure factors. Since tests showed that CC_ρ was comparable in selectivity to CC_R, we to decided to examine the CC based on complex values of $F_{\mathbf{h}} = A_{\mathbf{h}} + iB_{\mathbf{h}}$ and $X_{\mathbf{h}} = \alpha_{\mathbf{h}} + i\beta_{\mathbf{h}}$, where $X^*_{\mathbf{h}} = X_{-\mathbf{h}}$,

$$\mathrm{CC}_C = \frac{(\langle F_{\mathbf{h}}X^*_{\mathbf{h}}\rangle - \langle F_{\mathbf{h}}\rangle\langle X_{\mathbf{h}}\rangle)}{[(\langle|F_{\mathbf{h}}|^2\rangle - \langle F_{\mathbf{h}}\rangle^2)(\langle|X_{\mathbf{h}}|^2\rangle - \langle X_{\mathbf{h}}\rangle^2)]^{1/2}}$$
$$= \frac{(\langle A_{\mathbf{h}}\alpha_{\mathbf{h}} + B_{\mathbf{h}}\beta_{\mathbf{h}}\rangle - \langle A_{\mathbf{h}}\rangle\langle\alpha_{\mathbf{h}}\rangle)}{[(\langle|F_{\mathbf{h}}|^2\rangle - \langle A_{\mathbf{h}}\rangle^2)(\langle|X_{\mathbf{h}}|^2\rangle - \langle\alpha_{\mathbf{h}}\rangle^2)]^{1/2}}. \quad (6)$$

A problem arises in comparing maps that use phase indications that are weights of several different probable values (Bokhoven *et al.*, 1951; Blow & Rossmann, 1961). In such calculations, an average phase of $(\varphi_1 + \varphi_2)/2$ and an amplitude of $|F_{\mathbf{h}}|\cos[(\varphi_1 - \varphi_2)/2]$ are used. Given two maps, one for which $F_{\mathbf{h}}$ has either the phase φ_1 or φ_2 and the other for which $|F_{\mathbf{h}}|\cos(\varphi_1 - \varphi_2)$ has the phase-averaged value, the map variances are not identical,

$$\sigma^2(\rho)_{\varphi_1} = \sigma^2(\rho)_{\varphi_2}$$
$$= \sigma^2(\rho)_{\varphi_1+\varphi_2} \quad (7)$$
$$+ \ [1 - \cos^2(\varphi_1 + \varphi_2)]|F_{\mathbf{h}}|^2/V^2.$$

Small differences in the threshold values at which maps are modified will markedly effect the CC agreement between $F_{\mathbf{h}}$ and the $X_{\mathbf{h}}$. If a map in which a particular $F_{\mathbf{h}}$ is averaged is compared with one in which $F_{\mathbf{h}}$ is assigned a discrete phase value and the maps are compared at the same absolute threshold value for low-density elimination, the phase-averaged map will tend to compute a larger CC value, regardless as to whether the discrete tested value is closer to its true value or not. One can, however, avoid this impasse if one initially assigns phase-averaged values for all data, but for each particular $F_{\mathbf{h}}$ computes two maps, one for each discrete phase possibility. Under these circumstances the compared maps will have the same variance, but the comparison will require three FFTs for each SIR/SAS phase choice (a total of six) compared with only three FFTs per phase for the older method (since the initial starting map has already sampled one of the two permitted phase choices).

Table 1
Top 100 sign indications for the 3.75 Å cytochrome c_{550} SIR data.

The phases *A* were randomly selected as either the correct (+) or incorrect (−) Argand diagram choice as labeled following the serial number in columns 2 and 4. Correlation coefficient results based on both the amplitudes (CC_R) and complex values (CC_C) of $F_{\mathbf{h}}$ and $X_{\mathbf{h}}$ are listed. The results are sorted in descending order on the value of δCC followed by an asterisk if the entry indicates a phasing error. In *B*, phases were assigned the weighted average of the two SIR choices. The |δCC| results are sorted as in *A* and appended with an asterisk should the wrong phase choice be selected by the larger of the two CC values.

	A				*B*			
Rank	Ser	$\delta\mathrm{CC}_R$	Ser	$\delta\mathrm{CC}_C$	Ser	$\lvert\delta\mathrm{CC}_R\rvert$	Ser	$\lvert\delta\mathrm{CC}_C\rvert$
1	3 −	0.01390	23 −	0.01776	1	0.04523	1	0.03443
2	57 −	0.01378	4 −	0.01504	4	0.02713	5	0.03377
3	13 −	0.01108	5 −	0.01163	7	0.02401	7	0.03049
4	51 −	0.00956	19 −	0.00920	72	0.02325	4	0.02971
5	45 −	0.00952	54 −	0.00909	5	0.02272	15	0.02690
6	4 −	0.00908	72 −	0.00833	25	0.02187	23	0.02142
7	72 −	0.00890	45 −	0.00753	83	0.01925	14	0.02137
8	31 +	0.00636*	13 −	0.00694	24	0.01916	3	0.02027
9	83 −	0.00551	11 −	0.00681	19	0.01865	27	0.01996
10	46 +	0.00511*	53 +	0.00672*	63	0.01854	59	0.01815
11	80 −	0.00436	26 −	0.00662	85	0.01850*	72	0.01773
12	88 −	0.00331	3 −	0.00620	6	0.01800	9	0.01581
13	54 −	0.00321	51 −	0.00557	82	0.01706	25	0.01480
14	86 +	0.00290*	84 −	0.00549	15	0.01644	81	0.01326
15	33 −	0.00224	96 −	0.00542	14	0.01555	26	0.01297
16	52 +	0.00194*	25 −	0.00521	59	0.01530	92	0.01281
17	65 −	0.00180	83 −	0.00498	8	0.01510*	83	0.01265
18	2 −	0.00142	27 −	0.00494	23	0.01483	19	0.01264
19	60 −	0.00131	17 +	0.00408*	41	0.01449	94	0.01230
20	48 −	0.00121	63 −	0.00374	58	0.01430	32	0.01192
21	67 +	0.00108*	74 −	0.00366	27	0.01313	55	0.01191
22	76 −	0.00045	57 −	0.00320	92	0.01217	10	0.01188
23	23 −	0.00037	8 −	0.00292	32	0.01154	52	0.01152
24	78 +	0.00024*	97 +	0.00278*	45	0.01039	78	0.01113
25	55 +	0.00021*	58 −	0.00265	93	0.01037	58	0.01095
26	77 −	0.00014	86 +	0.00257*	60	0.01009	51	0.01032
27	100 +	−0.00003	9 −	0.00251	35	0.00981	48	0.01019
28	89 +	−0.00005	98 +	0.00209*	81	0.00979	82	0.01006
29	94 +	−0.00011	61 −	0.00177	57	0.00977*	90	0.00992
30	11 −	−0.00012*	16 +	0.00169*	48	0.00967	93	0.00976
31	66 −	−0.00029*	44 +	0.00166*	94	0.00965	45	0.00948
32	35 −	−0.00036*	52 +	0.00126*	51	0.00934	65	0.00921
33	12 +	−0.00041	47 −	0.00099	54	0.00920	41	0.00910
34	20 +	−0.00047	35 −	0.00097	52	0.00913	47	0.00852
35	91 +	−0.00053	65 −	0.00089	10	0.00893	98	0.00848
36	32 −	−0.00071*	33 −	0.00067	31	0.00884*	17	0.00838
37	61 −	−0.00087*	80 −	0.00006	40	0.00854	6	0.00836
38	58 −	−0.00087*	76 −	−0.00005*	80	0.00798	12	0.00786
39	19 −	−0.00102*	60 −	−0.00019*	74	0.00776*	39	0.00775
40	49 +	−0.00108	79 −	−0.00056*	12	0.00771	61	0.00769
41	50 +	−0.00120	90 −	−0.00061*	34	0.00743	64	0.00765
42	79 −	−0.00156*	71 −	−0.00080*	90	0.00741	60	0.00759
43	25 −	−0.00160*	68 +	−0.00085	47	0.00733	96	0.00740
44	44 +	−0.00164	69 +	−0.00092	11	0.00722*	99	0.00726
45	70 +	−0.00169	24 +	−0.00101	3	0.00720	16	0.00722
46	47 −	−0.00174*	38 +	−0.00121	65	0.00712	24	0.00704
47	40 −	−0.00180*	56 +	−0.00131	64	0.00670	38	0.00676
48	95 −	−0.00184*	67 +	−0.00151	26	0.00669	43	0.00667
49	71 −	−0.00200*	40 −	−0.00152*	70	0.00630*	56	0.00647
50	62 +	−0.00202	92 +	−0.00154	2	0.00630	67	0.00625
51	68 +	−0.00208	36 −	−0.00155*	20	0.00584	80	0.00609
52	56 +	−0.00223	100 +	−0.00162	46	0.00580*	76	0.00576
53	27 −	−0.00226*	50 +	−0.00202	17	0.00571*	84	0.00558
54	30 −	−0.00272*	88	−0.00225*	68	0.00568	71	0.00552
55	9 −	−0.00293*	85 +	−0.00304	43	0.00548	53	0.00532
56	75 +	−0.00305	87 +	−0.00313	88	0.00541	33	0.00501
57	16 +	−0.00307	32 −	−0.00313*	66	0.00534	89	0.00500
58	97 +	−0.00322	31 +	−0.00315	95	0.00530	63	0.00499
59	99 +	−0.00326	22 −	−0.00322*	56	0.00516	49	0.00499
60	15 +	−0.00333	70 +	−0.00470	79	0.00510	18	0.00484
61	73 +	−0.00346	37 +	−0.00489	18	0.00503*	29	0.00441
62	53 +	−0.00375	89 +	−0.00495	78	0.00489	37	0.00420
63	85 +	−0.00422	46 +	−0.00502	29	0.00485	2	0.00379
64	96 −	−0.00436*	28 −	−0.00523*	37	0.00473*	22	0.00370
65	5 −	−0.00436*	75 +	−0.00525	77	0.00471	85	0.00363*
66	43 +	−0.00440	93 +	−0.00544	22	0.00447*	54	0.00354
67	74 −	−0.00453*	64 −	−0.00560*	87	0.00444	42	0.00352
68	39 +	−0.00457	91 +	−0.00582	33	0.00430	97	0.00342
69	21 +	−0.00525	29 +	−0.00606	16	0.00414*	35	0.00335
70	84 −	−0.00542*	48 −	−0.00621*	89	0.00391	74	0.00334*
71	90 −	−0.00565*	82 +	−0.00635	21	0.00373	34	0.00333

Table 1 (continued)

	A				*B*			
Rank	Ser	δCC_R	Ser	δCC_C	Ser	$\lvert\delta CC_R\rvert$	Ser	$\lvert\delta CC_C\rvert$
72	28 −	−0.00571*	95 −	−0.00643*	98	0.00344	86	0.00329
73	10 −	−0.00581*	77 −	−0.00648*	28	0.00335	40	0.00294
74	93 +	−0.00588	73 +	−0.00652	53	0.00306	21	0.00262
75	36 −	−0.00596*	21 +	−0.00659	71	0.00281*	68	0.00251
76	17 +	−0.00609	81 +	−0.00719	61	0.00266	70	0.00235
77	98 +	−0.00610	49 +	−0.00719	99	0.00265	13	0.00228
78	38 +	−0.00625	43 +	−0.00739	38	0.00265	31	0.00206
79	37 +	−0.00633	62 +	−0.00745	44	0.00257	95	0.00202
80	22 −	−0.00648*	10 −	−0.00761*	91	0.00250*	46	0.00197
81	29 +	−0.00657	20 +	−0.00822	55	0.00227	87	0.00192
82	18 −	−0.00672*	99 +	−0.00838	67	0.00221	30	0.00162*
83	41 +	−0.00674	55 +	−0.00843	39	0.00214	62	0.00154
84	63 −	−0.00681*	94 +	−0.00855	69	0.00212	100	0.00150
85	81 +	−0.00720	66 −	−0.00908*	97	0.00210*	79	0.00150
86	59 +	−0.00758	39 +	−0.00921	86	0.00164*	11	0.00149*
87	69 +	−0.00829	59 +	−0.00965	13	0.00162	69	0.00096
88	87 +	−0.00884	15 +	−0.00976	76	0.00153*	8	0.00091
89	34 −	−0.00919*	78 +	−0.00981	100	0.00133*	36	0.00083
90	42 +	−0.01030	30 −	−0.00996*	49	0.00113*	77	0.00079*
91	82 +	−0.01040	41 +	−0.01047	62	0.00105*	91	0.00066*
92	92 +	−0.01065	14 +	−0.01435	73	0.00101	88	0.00062
93	26 −	−0.01069*	12 +	−0.01449	30	0.00100*	28	0.00043*
94	14 +	−0.01141	18 −	−0.01564*	36	0.00088	66	0.00033
95	64 −	−0.01239*	42 +	−0.01732	84	0.00086	73	0.00032*
96	1 −	−0.01246*	7 +	−0.01863	9	0.00081*	75	0.00031
97	24 +	−0.01511	34 −	−0.02011*	42	0.00033	44	0.00029*
98	7 +	−0.01518	6 +	−0.02014	75	0.00031	50	0.00023*
99	8 −	−0.01566*	2 −	−0.02484*	50	0.00017	57	0.00011*
100	6 +	−0.01568	1 −	−0.02664*	96	0.00001	20	0.00004*

2. Test example

Error-free SIR data for cytochrome c_{550} were computed from the published coordinates (PDB entry 155c; Timkovich & Dickerson, 1976) as reported previously. The iron-containing protein was considered to be the heavy-atom derivative; the native structure omitted the Fe-atom site. The primary data file contained h, k, l, $|F|$, φ_1, φ_2 and $|F\sin[(\varphi_1 - \varphi_2)/2]|$ and was sorted in decreasing order of the magnitude of the $|F\sin|$ term. This places those reflections at the top of the list that have the greatest effect on altering map features upon changing phase values from φ_1 to φ_2. Our previous tests had compared both SIR and SAS data and results for three different resolution ranges: 2.65, 3.75 and 5.3 Å. In this presentation, it will suffice to present only SIR results for the 3.75 Å range (1299 reflections). In Table 1, we compare the old scheme (*A*) in which the phases of the initial map were randomly chosen as either φ_1 or φ_2 [r.m.s.$\delta(\varphi)$ = 64°] and the new scheme (*B*) for which all data were initially assigned the averaged phase and amplitude [r.m.s.$\delta(\varphi)$ = 45°]. Results based on both CC_R and CC_C for the top 100 SIR phase choices of the sorted data list are presented. Columns were sorted in decreasing order on the values of $\delta CC = CC_{\varphi 1} - CC_{\varphi 2}$. Phase choices that have the largest δCC values tend to be more reliable and sort to the top of each column.

3. Summary

The results in part *A* of Table 1 clearly show a small advantage in using CC_C in preference to CC_R in resolving the SIR phase ambiguity. In column 3 there are seven phasing errors (*) noted in the top 25 indications based on δCC_R, compared with three errors in column 5 as indicated by δCC_C. However, the results from part *B* of Table 1 are even more encouraging. Whereas the first errors occur at lines 11, 17 and 29 in column 7 for CC_R, it is remarkable to note that CC_C offers far more selectivity in identifying the correct SIR phase choice based on the same basis set of $F_{\mathbf{h}}$ and $X_{\mathbf{h}}$ values. The first SIR phasing error occurs at line 65 in column 9, compared with line 11 in column 7. Given these results, it is fairly easy to reduce the overall phase error of the 3.75 Å SIR data set to less than 10° in a small number of passes through the full set of data.

4. Closing remarks

We previously noted that our concept of unbiased $X_{\mathbf{h}}$ estimates is the reciprocal-space analogue of the well known 'omit-map' procedure (Bhat & Cohen, 1984) that is performed in real space. Figures of merit, such as the free R value (Brünger, 1993), which rely on excluding small fixed or rotating (Roberts & Brünger, 1995; Cowtan & Main, 1996) subsets of data from the refinement calculations and using their extrapolated values to monitor the correctness of the refinement model, can in principle be reformulated with $X_{\mathbf{h}}$ estimates to end the need to exclude data from the refinement process. A reviewer has called attention to similarities between (1) and an analysis of the 'solvent-flipping' scheme used in density modification (Abrahams, 1997). The subtraction of a 'γ-correction' from the solvent mask can ensure that its zero-order Fourier transform $G_0 = 0$. This would in effect correct the map-extrapolated values of $F_{\mathbf{h}}$ for the $F_{\mathbf{h}}G_0$ term, but the contribution of $F_{-\mathbf{h}}G_{2\mathbf{h}}$ and other symmetry-related terms, $F_{\mathbf{h}_j}G_{\mathbf{h}-\mathbf{h}_j}$, would remain neglected.

Given that CC_C offers greater selectivity than CC_R in the applications described, one might expect that the calculation of likelihood [Bricogne, 1984, equation (4.16)] could be modified to consider the degree of phase agreement between our initial $F_{\mathbf{h}}$ and its unbiased extrapolated 'non-basis set' value. The argument of the exponential term used to compute the likelihood is normally of the form $\exp[-N(|U_{\mathbf{h}}^{\rm obs}|^2 + |U_{\mathbf{h}}^{\rm ME}|^2)]$. One might expect that modifying this expression to something like $\exp[-2N|U_{\mathbf{h}}^{\rm obs}U_{\mathbf{h}}^{\rm ME}| \times \cos(\varphi_{\mathbf{h}} - \varphi_{\mathbf{h}}^{\rm ME})]$ might add a significant degree of selectivity to this measure in the context of the calculations described in this paper. However, whether this selectivity is any better than that provided by CC_C remains to be demonstrated.

We gratefully acknowledge financial support received through NIH grant GM-46733.

References

Abrahams, J. P. (1997). *Acta Cryst.* D**53**, 371–376.
Bhat, T. N. & Cohen, G. H. (1984). *J. Appl. Cryst.* **17**, 244–248.
Blow, D. M. & Rossmann, M. G. (1961). *Acta Cryst.* **14**, 1195–1202.
Bokhoven, C., Schoone, J. C. & Bijvoet, J. M. (1951). *Acta Cryst.* **4**, 275–280.
Bricogne, G. (1984). *Acta Cryst.* A**40**, 410–445.
Brünger, A. T. (1993). *Acta Cryst.* D**49**, 24–36.
Cowtan, K. D. & Main, P. (1996). *Acta Cryst.* D**52**, 43–48.
Langs, D. A., Blessing, R. H. & Guo, D. Y. (2001). *Acta Cryst.* D**57**, 574–578.
Read, R. J. (1986). *Acta Cryst.* A**42**, 140–149.
Read, R. J. (1997). *Methods Enzymol.* **277**, 110–128.
Roberts, A. L. U. & Brünger, A. T. (1995). *Acta Cryst.* D**51**, 990–1002.
Timkovich, R. & Dickerson, R. E. (1976). *J. Biochem.* **251**, 4033–4046.