Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Critical initial real-space refinement in the structure determination of arginine kinase

Genfa Zhou,[a]† Thayumanasamy Somasundaram,[a] Eric Blanc,[a]‡ Zhi Chen[a] and Michael S. Chapman[a,b]*

[a]Institute of Molecular Biophysics, Florida State University, Tallahassee, FL 32306-4390, USA, and [b]Department of Chemistry, Florida State University, Tallahassee, FL 32306-4380, USA

† Present address: Enders 673, Howard Hughes Medical Institute, 320 Longwood Avenue, Boston, MA 02115, USA.
‡ Present address: Global Phasing Ltd, Sheraton House, Castle Hill Business Park, Cambridge CB3 0AX, England.

Correspondence e-mail: chapman@sb.fsu.edu

Received 17 August 1998
Accepted 18 January 1999

Arginine kinase (AK), a homologue of creatine kinase, catalyses the reversible transfer of a phosphoryl group between a guanidino phosphate and ADP. The family of phosphagen kinases eluded structure determination for over 25 years until an inactive form creatine kinase (CK) structure was determined [Fritz-Wolf *et al.* (1996). *Nature (London)*, **381**, 341–345]. The structure determination of the active-form transition-state complex was non-trivial, owing to the distant relatedness and domain reorientation of AK compared with CK. Phases from a molecular-replacement solution of the large domain, supplemented by single isomorphous replacement and inter-crystal averaging, did not reveal interpretable electron density for the small domain. Reciprocal-space refinement of the initial model ($R^{\rm free}$ = 0.54) by any of the commonly used methods, including *post facto* application of maximum-likelihood methods, led to overfitting without significant improvement of the partial initial model. By contrast, in the local real-space refinements which proved successful, the interdependence of atoms is limited to immediate neighbors, and atomic positions are not influenced by errors or omissions in remote parts of the structure. Modest improvement was possible without overfitting, and this was critical to the calculation of improved phases. Phases were refined and extended from 4.0 to 2.5 Å resolution by Fourier inversion of omit maps, combination with isomorphous replacement phases and averaging between crystal forms, after several batches of real- and reciprocal-space atomic refinement. The final structure refinement, against a 1.86 Å cryo data set yielded a high-quality model with R = 0.196 and $R^{\rm free}$ = 0.224.

1. Introduction

In recent years, there have been substantial methodological advances in experimental phase determination, model building and refinement (Brünger *et al.*, 1997; for example, Hendrickson & Ogata, 1997; Jones & Kjeldgaard, 1997). Nevertheless, progressing from a crude, perhaps partial, atomic model which can be built into an initial map that may be very poor, to one which will refine towards the correct structure remains a challenge that can delay macromolecular structure determination for many years. Often, the crude models have little correlation with the diffraction data, with R factors sometimes exceeding 0.5. Such models are susceptible to overfitting during many types of reciprocal-space refinement, during which it may be possible to improve the agreement with diffraction data and lower the working R factor ($R^{\rm work}$), without a real improvement in the model. It is only recently that the methods of cross validation which reveal overfitting have been in common use (Brünger, 1997*a*). Thus,

Table 1
Crystallographic statistics.

Data collection	Native	Native	Native	K_2PtCl_4	Thiomersal
Crystal form	1	2	2	2	2
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
a (Å)	64.2	66.4	65.4	66.7	66.8
b (Å)	66.0	72.5	70.9	72.8	72.6
c (Å)	86.7	80.4	80.4	80.3	80.4
Resolution limits (Å)	100–2.25	100–2.25	100–1.86	100–2.5	100–3.0
Number of reflections	17562	18378	35225	13902	7103
Completeness (at resolution limit) (%)	94.6 (88.3)	94.3 (76.5)	98.9 (98.5)	98.9 (97.3)	84.5 (77.5)
Redundancy	3.3	5.6	4.5	4.4	4.9
$\langle I/\sigma_I \rangle$ (at resolution limit)	18.5 (4.6)	33.1 (4.6)	37.2 (12.2)	35.0 (6.3)	24.9 (10.1)
$R_{\rm merge}$† (at resolution limit) (%)	7.0 (26.5)	6.0 (31.1)	4.7 (15.7)	4.8 (24.1)	7.5 (17.5)

† $R_{\rm merge} = \sum |I_h - \langle I \rangle| / \sum |I_h|$, where $\langle I_h \rangle$ is the average intensity of symmetry-equivalent observations.

it is becoming apparent that problems previously blamed on the intermediate stages of refinement (with $R^{\rm work} \simeq 0.4$) are likely to have their origins in the failure of initial refinement to achieve any real improvement of the model. A particularly intractable class of problems are those structures for which it is initially only possible to build a partial atomic model. Here, through application to a 'problem' structure, it is demonstrated that real-space refinement is particularly appropriate for this initial refinement, owing to its wide convergence radius and ability to achieve real improvements in a poor model without overfitting.

Arginine kinase (AK) is one of the phosphagen kinase family, including creatine kinase (CK), which catalyze the reversible transfer of a phosphoryl group between guanidino phosphate compounds and ADP. Owing to their central role in buffering cellular ATP energy levels and their use as a paradigm of enzymology, extensive biochemical studies date back 40 years (Kenyon & Reed, 1983). During the last 25 years, there have been at least eight independent preliminary crystallographic reports for these enzymes (*e.g.* McPherson, 1973). Not for lack of trying, it was not until two years ago that the challenges of atomic structure determination were first overcome with an octameric mitochondrial CK (Mi_b-CK; Fritz-Wolf *et al.*, 1996). Two 3 Å structures were determined of the enzyme in an inactive open configuration (Forstner *et al.*, 1996) either in the apo form or in complex with one of the substrates, ATP. The structures revealed the path of the polypeptide backbone, but not the details of the active site, which was largely disordered.

Here, the methods used in the 1.86 Å resolution structure determination of the monomeric arginine kinase from *Limulus polyphemus* (horseshoe crab; Zhou *et al.*, 1998) are reported. It was overexpressed in *Escherichia coli* (Strong & Ellington, 1996), purified and co-crystallized in the closed active configuration (Forstner *et al.*, 1996) as a transition-state-analog complex (TSAC) with arginine, MgADP and nitrate (mimicking a planar phosphoryl group during transfer; Zhou *et al.*, 1997). The structure revealed the amino acids critical to catalysis and indicated a reaction mechanism which involved the precise pre-alignment of the two substrates so that the orbitals were aligned for optimal reaction trajectories (Zhou *et al.*, 1998). Crucial to successful structure determination was the initial atomic refinement in real space. Real-space refinement implicitly uses phase information, because the electron-density maps against which the structure is refined are calculated with the phases. The use of real-space methods runs counter to the conventional wisdom that inaccurate phases should be discarded at the earliest opportunity. On the other hand, given the effort expended in model fitting (Jones & Kjeldgaard, 1997), a common-sense argument could be made that the (real-space) fit of model to map should be optimized before other types of atomic refinement. The method which was used (*RSRef*; Chapman, 1995) is conceptually similar to prior-refinement methods (Diamond, 1971; Jones & Liljas, 1984) and to the *RSR* routines of molecular modeling (Jones *et al.*, 1991), with a few important exceptions. Firstly, the resolution limits are explicitly incorporated into the calculation of model electron density. Secondly, full bonded and non-bonded stereochemical restraints are imposed (using routines from *TNT*; Tronrud *et al.*, 1987) and, thirdly, it is a restrained gradient-descent method in which fit to the map and stereochemistry are handled simultaneously.

Prior studies had suggested that real-space methods might be of use in marginal initial refinements: (i) pre-refinement against even a poor multiple isomorphous replacement map improved subsequent reciprocal-space refinement (Chapman & Blanc, 1997) and (ii) real-space refinement was a critical component in completing atomic models of tryptophanyl tRNA synthetases (Carter *et al.*, 1996).

Here, the role of initial real-space refinement in the solution of difficult crystal structures is illustrated with a case study: arginine kinase. The context is set with a description of the first parts of the structure determination which led to a crude initial model which was not refinable by available reciprocal-space methods. Attempts to refine in real and reciprocal space are compared. The value of the real-space methods is emphasized with a description of the ease with which an accurate high-resolution structure was subsequently obtained. Finally, a more rigorous *post facto* comparison of the early steps of arginine kinase refinement is made, with tests including methods which became available during our structure determination – torsion-angle refinement (Rice & Brünger, 1994) and maximum-likelihood techniques (Murshudov *et al.*, 1997; Pannu & Read, 1996).

2. Methods and results of the arginine kinase structure determination

2.1. From crystals to partially interpretable map

2.1.1. Native data collection. Diffraction data were collected using 1.54 Å Cu $K\alpha$ X-rays from a Rigaku RU-200

Table 2
Statistics for rotation (RF) and translation function (TF).

Molecular-replacement probe	Crystal form	Calculation	Highest peak heights†‡	Retrospective analysis of correct solution		
				Peak rank†	Peak height†‡	Next peak†‡
CK whole subunit	1	RF	14.2/0.055	4/1	13.5/0.055	—/0.052
CK whole subunit	1	TF	12.6/failed	1/none	12.6/failed	10.2/failed
CK whole subunit	2	RF	13.4/—	11/—	10.9/—	—/—
CK whole subunit	2	TF	12.0/—	1/—	12.0/—	11.6/—
CK large domain	1	RF	16.2/0.059	1/2	16.2/0.058	14.3/0.052
CK large domain	1	TF	16.1/0.18	1/1	16.1/0.18	11.5/0.14
CK large domain	2	RF§	14.0/0.064	1/1	14.0/0.064	13.4/0.048
CK large domain	2	TF	16.5/—	1/—	16.5/—	14.8/—
AK form 1 data	**2**	**RF**	**40.7/—**	**1/—**	**40.7/—**	**18.9/—**
CK small domain	1	RF	19.4/0.061	5/1	15.9/0.061	—/0.046
CK small domain	1	TF	24.7/0.061	1/1	24.7/0.061	21.2/—
CK small domain	2	RF	failed/failed	>50/>50	failed/failed	failed/failed

† Scores from two programs are given in the format *AMoRe*/*X-PLOR*. The 'correct' solution is the one which is consistent between the two crystal forms and with the isomorphous replacement data. ‡ For *AMoRe*, rotation-function peak heights are given in units relative to the estimated standard deviation and translation functions are correlation coefficients × 100; for *X-PLOR*, correlation coefficients are given. § Search limited to ±20° about the whole subunit orientation.

HB generator and a 0.3 mm collimator. 15 min exposures were recorded with an R-AXIS II image-plate system 90 mm from the crystal. Data were collected at room temperature from two $P2_12_12_1$ crystal forms grown under identical conditions which diffracted to 2.25 Å resolution (Table 1). There were indications of modest radiation damage. At the end of the form 2 data collection, after 150° rotation and 19 h of exposure, the scaling constant was reduced by 13% and the *B* factor had increased to 5.7 Å^2 (corresponding to a 30% drop of intensity at 2.5 Å resolution). A third native data set was collected at 1.86 Å resolution from form 2 crystals mounted in 0.7–1.0 mm rayon loops and cooled to 100 K in a nitrogen stream (Oxford Cryosystems, Inc.) after cryo-protection in 24.5%(*w*/*v*) glycerol (Rodgers, 1994). There was no evidence from the scaling of any radiation damage (*B* changed by 1 Å^2). Data were processed with *BioTex* (Molecular Structure Corp.) and *HKL* (Otwinowski & Minor, 1997).

2.1.2. Molecular replacement. The molecular-replacement solution was marginal for reasons that, in retrospect, are now clear. There is a domain rearrangement upon substrate (or transition-state analog) binding (Zhou *et al.*, 1998). The search probe was the Mi_b-CK subunit A structure (Fritz-Wolf *et al.*, 1996), determined as an apo-enzyme at 3 Å resolution. The large domain had modest sequence identity (43%) and nominally accounted for 70% of the structure, but parts of the probe structure were was highly disordered in the absence of substrates (Fritz-Wolf *et al.*, 1996). The small domain accounted for 26% of the amino acids with (only) 28% sequence identity.

Cross-rotation functions (Navaza & Saludjian, 1997; Rossmann, 1972) were calculated between CK models and AK (room temperature) data for both crystal forms. The eight Mi_b-CK subunits have different loop structures, but all worked equally well. Attempts were made with a single (two-domain) probe and also with separate large and small domains to allow for the rearrangement. Rotation and translation functions were calculated using two different implementations in *AMoRe* (Navaza, 1994) and *X-PLOR* (Brünger, 1997*b*). Results are shown (Table 2) for calculations using data between 8 and 4 Å or 10 and 4 Å with an $F/\sigma(F)$ cutoff of 2 in *X-PLOR*, an integration radius for rotation functions of 20 Å and a step size of 1° (*AMoRe*) or 2.5° (*X-PLOR*). These results were cross-checked with calculations (not shown) with altered high-resolution limit (3–6 Å), low-resolution limit (6–20 Å) and integration radius (10–22 Å). The larger C-terminal domain (Mi_b-CK residues 113–363) was oriented first and the N-terminal domain (1–95) was then added. With *AMoRe*, the top 50 solutions for the large domain were used for small-domain rotation-function searches. Computation was reduced in *X-PLOR* by assuming that the domain rearrangement in AK compared with Mi_b-CK would involve small domain rotations less than ±20° and translations of less than ±7 Å.

Molecular-replacement solutions were not strong (Table 2) and required verification. For the whole subunit probe, *AMoRe* and *X-PLOR* gave different sets of possible rotation-function solutions, although some were in common. The large domain gave clearer results, with both programs ranking the same solution highly. This solution was consistent with a strong peak in a cross-rotation function calculated between diffraction data of the two crystal forms (Table 2, bold), which showed that the two forms are related by a rotation of the protein by ∼160° about a vector approximately along (1,1,0). Confidence in the weak rotation-function solution would later be bolstered by its use in calculating phases which revealed heavy-atom sites whose locations had been determined independently (see below).

As for the small domain, reasonable rotation- and translation-function solutions were only found for crystal form 1. Several variants of the probe were tried, with a polyglycine model giving the clearest solutions. The *AMoRe* rotation function peak 5 gave the best translation function. During Patterson correlation refinement the orientation changed by ∼6°, closer to the top-ranked solution of a Patterson correlation search (Brünger, 1997*b*). In the unsuccessful attempts with crystal form 2, a Patterson subtraction method (Zhang & Matthews, 1994) was tried in addition to *AMoRe* and *X-PLOR*.

Table 3
Heavy-atom parameters.

K_2PtCl_4. Phasing power = 1.0; R_{iso} = 0.18.†

	Fractional coordinates				
Site	*X*	*Y*	*Z*	*B* (Å^2)	Occupancy
1	0.443	0.460	0.199	25	0.62
2	0.739	0.352	0.183	25	0.50
3	0.513	0.937	0.265	25	0.20

Thiomersal. Phasing power = 1.2; R_{iso} = 0.17.†

	Fractional coordinates				
Site	*X*	*Y*	*Z*	*B* (Å^2)	Occupancy
1	0.441	0.463	0.185	25	0.72
2	0.850	0.933	0.086	25	0.37
3	0.514	0.651	0.117	25	0.41

† $R_{iso} = \sum |F_{PH} - F_P| / \sum |F_P|$, or the mean fraction isomorphous difference, and phasing power = f_H/E, where f_H is the heavy-atom structure amplitude and E is the lack-of-closure error.

Retrospectively, alignment of the CK structure to the completed AK structure as separate rigid-body domains showed that the original molecular-replacement solutions deviated from the optimal superimposition with r.m.s. coordinate errors of 0.37 and 1.76 Å for the large and small domains, respectively. Retrospective analysis also showed that, relative to the apo-form, the small domain of the substrate-bound AK form was rotated 13° with negligible translation. Thus, the weakness of the molecular-replacement solution stemmed from the difficulty of precisely defining the orientation of a domain which accounted for 26% of the structure, and was only distantly related to the phasing probe (28% sequence identity).

2.1.3. Multiple isomorphous replacement. Heavy-atom derivatives would be needed to verify the marginal molecular-replacement solution and to improve inadequate phases. Two derivatives were found by soaking form 2 crystals in TSAC buffer containing 1 m*M* K_2PtCl_4 or 0.5 m*M* thiomersal for 12 h. Diffraction data were collected at room temperature to the 3 Å resolution limit (Table 1). The first platinum site was determined by inspection of the Harker sections of the isomorphous difference Patterson map (Stout & Jensen, 1989). Additional sites were determined from an isomorphous difference Fourier synthesis (Ramachandran & Srinivasan, 1970) phased from the first site. These additional sites were consistent with self- and cross-vectors in the Patterson Harker sections and with an automated Patterson search using *RSPS* (Knight, 1989; Terwilliger *et al.*, 1987). The thiomersal binding sites were determined from a difference Fourier phased with the K_2PtCl_4 derivative and were consistent with a difference Patterson. Heavy-atom parameters (Table 3) were refined by maximum-likelihood methods using *MLPHARE* (Otwinowski, 1991).

2.1.4. Validation of the molecular-replacement solution. Consistency between the molecular-replacement solution and the heavy-atom data was checked by looking for the heavy-atom sites in isomorphous heavy-atom difference Fourier maps. Phasing with the appropriately positioned large CK domain gave a difference map with the highest peaks at the Pt sites. The small domain gave a number of (similarly sized) peaks, of which three were the expected Pt sites. This reflected the weaker phasing power and greater uncertainty in position, but confirmed that the small domain added some phase information.

2.1.5. Phase combination, inter-crystal averaging and density modification. Isomorphous replacement and molecular-replacement phases for crystal form 2 were combined and coefficient weights were calculated by the σ_A procedure (Read, 1985). At the start of phase improvement, only the Pt derivative was available, and the small domain was omitted from the molecular-replacement phasing, so as not to bias future maps with an uncertain model. Crystal form 1 was phased using the large domain of the molecular-replacement

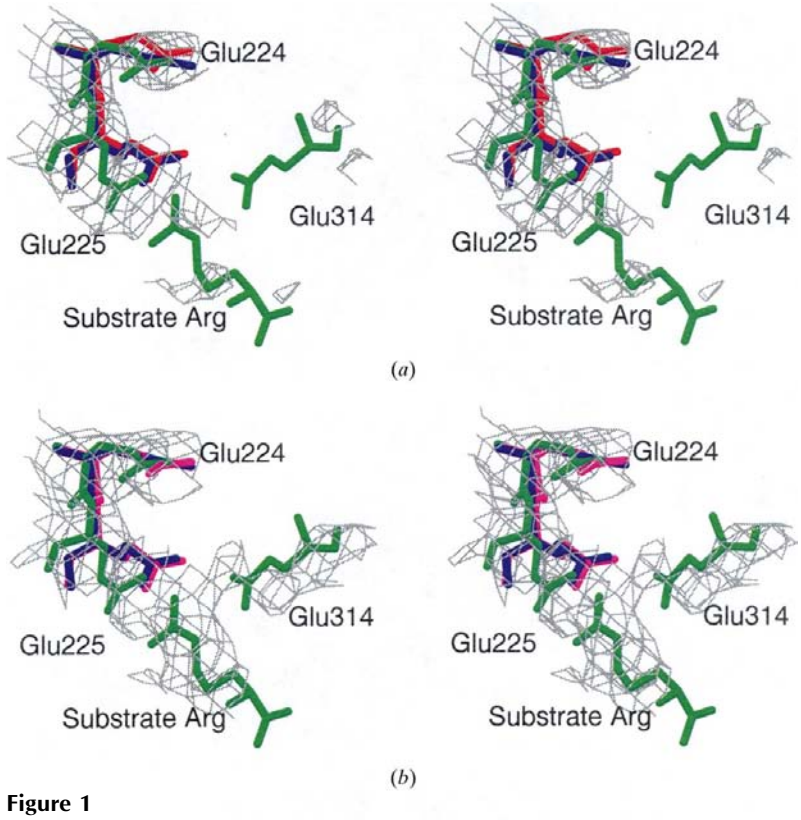

Figure 1
Quality of the electron density at several stages of structure determination. The stereoviews show the electron density contoured at the 1σ (*a* and *b*) or 2σ (*c* and *d*) level for the substrate arginine and some of the surrounding active-site amino acids. (*a*) The initial 3 Å map is shown with the molecular-replacement phasing model 0 (R^{free} = 0.543) in red, rebuilt model 1 (R^{free} = 0.544) in blue and, as in all panels, the final structure in green. (*b*) The 3 Å map, calculated with the initial real-space refined model 2 (R^{free} = 0.518) reveals density for parts that were initially modeled completely incorrectly (Glu314) or were missing (substrate Arg). This map is shown with rebuilt model 1 in blue (R^{free} = 0.544) and the initial real-space refined model 2 (R^{free} = 0.518) in pink.

probe. The protein electron density was then averaged between the two crystal forms using *DMMulti*, a multi-crystal extension of *DM* (Cowtan & Main, 1993). As the small-domain rotation function had been unsuccessful in crystal form 2, the small and large domains were assumed to be in the same relative alignment as in form 1 crystals for the purposes of deriving an envelope for inter-crystal averaging.

The molecular envelope needed for averaging was defined using the program *MAMA* (Kleywegt & Jones, 1993), by masking pixels within 4 and 10 Å of large- and small-domain phasing model atoms, respectively (allowing for greater uncertainty in the small-domain model). Consistency between the two crystal forms was forced through iterations of transforming between the two crystal forms, expanding according to the crystallographic symmetry and trimming from the parts of the mask which overlapped most, starting with the small domain. After smoothing, the mask accounted for 85% of the form 2 unit cell.

Density, averaged at 4 Å resolution, was placed into the form 2 unit cell and Fourier inverted to calculate improved phases. Density outside the envelope was not modified. During 20 iterations, the resolution was extended from 4 to 3 Å. Density for the small domain had become stronger, so automatic methods were used to calculate a second mask specifically for solvent flattening (Wang, 1985), which was added for the next 20 cycles at 3 Å. The mask occupied 33% of the unit cell, compared with a calculated 48% solvent fraction. During phase refinement, the density-modification cross-validated *R* factor (Cowtan & Main, 1996) dropped from 74 to 51%, indicating successful phase refinement. The average figure of merit rose from 0.32 to 0.67 (form 1) and from 0.50 to 0.73 (form 2).

2.2. Critical use of real-space methods to initiate refinement of a crude model

2.2.1. Setting the stage. At this point, a critical hurdle was encountered. Reciprocal-space atomic refinement of the molecular-replacement probe against the AK diffraction amplitudes led only to increasing $R^{\rm free}$ (even though $R^{\rm work}$ fell to about 0.45). Doubts remained about the exact location of the small domain, because density in this region of the averaged map was uninterpretable and because rigid-body refinement of the phasing model gave equally good results whether the small domain was present or not. Following the 3 Å phase refinement, some of the differences between the CK and AK large domains became apparent in the electron-density map of crystal form 2. The map was not yet of high quality. It had been phased using a significantly different atomic model, and weak derivatives with phasing power of ∼1 were the strongest that had been found which were still isomorphous. These conditions are not uncommon and frequently block or delay structure determination. Progress would depend upon improving the phases through improving the model of the large domain until the small domain could be seen and modeled.

2.2.2. Attempts to refine the crude model. The 3 Å phase-refined map was of sufficient quality (Fig. 1*a*) to model (Jones *et al.*, 1991) much of the large-domain AK sequence, remodel some of the loops and improve the backbone conformation. In rebuilding, standard modeling methods (Jones *et al.*, 1991) were combined with local real-space refinement (Blanc *et al.*, 1998) in which small regions of the structure are least-squares refined. Rebuilding had little effect upon $R^{\rm free}$ (which changed from 0.543 to 0.544; quoted *R* factors are for crystal form 2). Phase refinement was repeated using the initial AK model instead of the CK phasing model. Reciprocal-space conjugate-gradient refinement using *TNT* (Tronrud *et al.*, 1987) or *X-PLOR* (Brünger *et al.*, 1987) failed, as indicated by marginally improved or worsening $R^{\rm free}$, even though $R^{\rm work}$ decreased by ∼10%; this was evidence of overfitting (Brünger, 1997*a*). The exact statistics depended on the weights used, but the best showed marginal improvement of $R^{\rm free}$ from 0.544 to 0.539. Refinements were tried with or without inclusion of a polyglycine molecular-replacement model for the small domain. Molecular-dynamics methods fared worse than conjugate-gradient methods, leading to greater overfitting.

In contrast to reciprocal-space attempts, real-space refinement of the entire large domain (Chapman, 1995) improved the $R^{\rm free}$ of the model

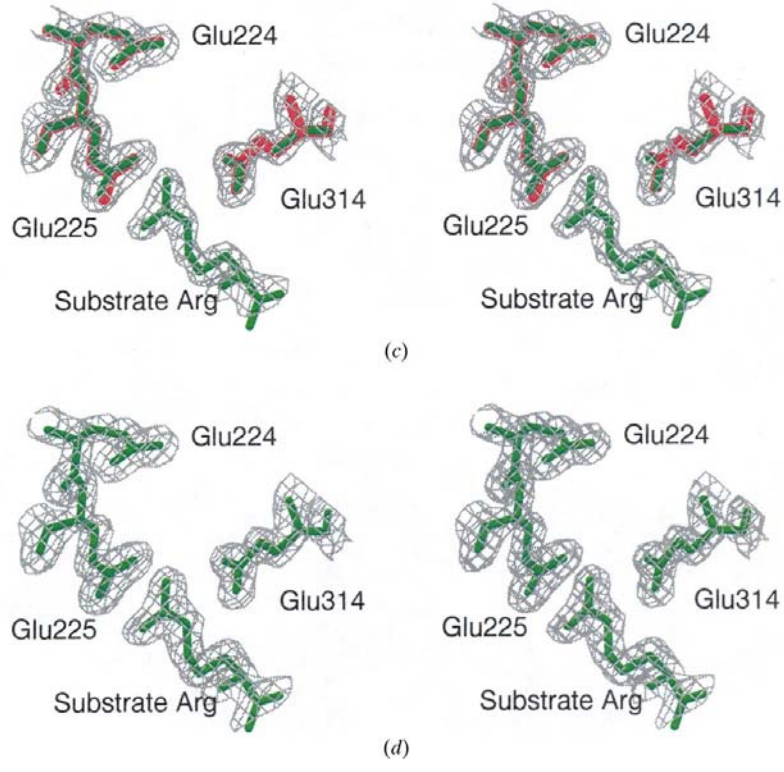

Figure 1 (continued)

(*c*) This is the 1.86 Å map used to model the substrates that had hitherto been omitted to avoid possible phase bias. For the model shown in brown, $R^{\rm free}$ is 0.287. (*d*) The final map at 1.86 Å resolution is shown with the final AK model in green.

Table 4
2.5 Å phase and model refinement with the room-temperature data.

Model	Obtained from	R^{work}	R^{free}†	Resolution (Å)	R.m.s. deviation from final structure (Å)		Phasing				
					Main chain	Side chain	Atomic model	Pt derivative	Hg derivative	Inter-crystal averaging	Solvent flattening
0	Molecular replacement	0.522	0.543	10–3.0	N/A	N/A	CK L‡				
1	Modeling + local real-space refinement	0.516	0.544	10–3.0	3.34	4.98	0 (CK L)	Yes	No	Yes	Yes
2	Real-space refinement	0.513	0.518	10–3.0	3.34	4.97					
3	Modeling	0.473	0.508	10–3.0	2.98	3.56	2 + CK S‡	Yes	Yes	Yes	Yes
4	Reciprocal-space refinement	0.447	0.493	10–3.0	2.90	3.05					
5	Modeling	0.444	0.460	10–3.0	2.87	2.84	4	Yes	Yes	Yes	Yes
6	Reciprocal-space refinement	0.395	0.396	10–3.0	1.19	2.08					
7	Modeling	0.390	0.402	10–3.0	1.19	1.97	6	Yes	Yes	Yes	Yes
8	Real-space refinement	0.342	0.385	10–3.0	1.15	1.90					
9	Reciprocal-space refinement	0.333	0.368	10–2.5	1.09	1.84					
10	Modeling	0.311	0.356	10–2.5	0.61	1.55	9	Yes	Yes	Yes	Yes
11	Real-space refinement	0.304	0.339	10–2.5	0.49	1.08					

† 3% of reflections were omitted from map calculation and refinement and used for cross-validation. *R* factors were calculated from data selected with an $F/\sigma(F)$ cutoff of 2.0. All quoted *R* factors are with respect to crystal form 2. Using data beyond 10 Å resolution, 7429 reflections were used for refinements to 3 Å, 12 608 to 2.5 Å and 15 890 to 2.3 Å, excluding the 242 390 and 493 reflections, respectively, used for R^{free} calculation. ‡ L is used for large domain, S for small.

from 0.544 to 0.518 (Fig. 1*b*, Fig. 2 and Table 4). The appearance of some density for unmodeled substrates (Fig. 1*b*) was another indication of progress. (Additional details of the real-space protocol will be provided in the *post facto* analysis.)

Conjugate-gradient and simulated-annealing reciprocal-space refinement of the real-space refined model again failed, leading only to overfitting and degradation of the model, with larger decreases in R^{work} than in R^{free}. In one attempt, a 0.04 (4%) decrease in R^{work} was accompanied by an increase in R^{free} to 0.519. A more aggressive attempt reduced R^{work} by 0.12 to ∼0.4, but led to larger increases in R^{free}. It would be only when R^{free} became ∼0.50 that reciprocal-space refinement of such a crude partial model would begin to converge. The model was brought within this convergence radius by recalculating a map from the phases of the real-space refined model, remodeling the large domain and adding a backbone model for the small domain.

More detailed comparisons of initial refinement attempts will be given in the *post facto* analysis.

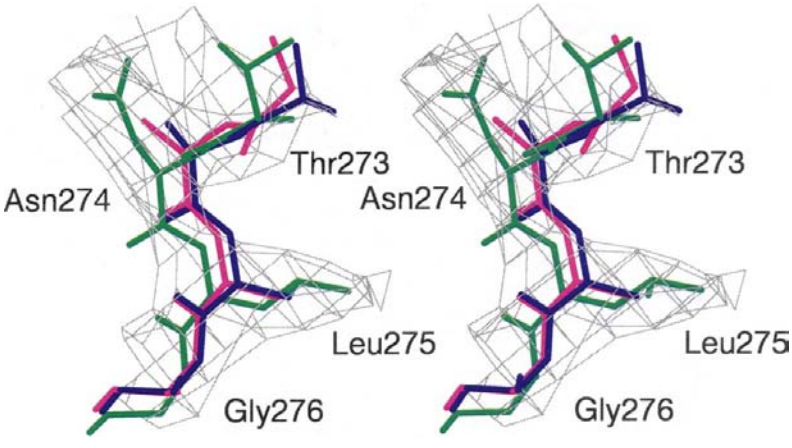

Figure 2
Model improvement during initial real-space refinement. This typical region shows that the model improvement from the blue model 0 (R^{free} = 0.544) to model 1 (pink; R^{free} = 0.518) is modest in comparison to the final model (green). Although it may seem obvious that the backbone of residues 274–5 should move farther, at this stage refinement was held back by errors at 273 and 276. The modest model improvement seen here started a process in which the map was improved substantially (Fig. 1) until it became apparent how to correct errors such as those at residues 273 and 276.

2.3. Completion of the high-resolution structure

Following the maxim that the proof of the pudding is in the eating, this section documents how the results of real-space refinement led to a successful completion of the structure.

2.3.1. 2.5 Å structure determination by iterative phase, map and model improvement. Once refinement progressed to $R^{free} \simeq 0.5$, it was possible to incorporate reciprocal-space refinement in iterative phase/model improvement. Phases from the second derivative (thiomersal) had become available and were incorporated at the start. Iterations involved re-refinement of the current phases by density modification and inter-crystal averaging, new map calculation, modeling and atomic refinement in both real and reciprocal space (Table 4). The resolution was extended gradually from 3.0 to 2.5 Å. For phase refinement, the molecular envelope was gradually reduced to mask only pixels within 3 Å of atoms and the solvent fraction was increased to 43%. Particular attention was paid to the following.

(i) To minimize the chance of phase bias towards an atomic model, the starting phases for an iteration were calculated by back-transformation of an omit map (Bhat & Cohen, 1984).

(ii) The transformation operators between the two crystal forms were refined independently for the small and large domains (using the procedure in *DMMulti* which maximizes the agreement between the electron density; Cowtan & Main, 1993).

Table 5
High-resolution refinement against cryo-diffraction data.

Batch	Process	R^{work}	Number of reflections	R^{free}†	Number of reflections	Resolution (Å)
0		0.486 (0.514)	7273	0.473 (0.472)	240	10–3.0
1	Rigid body (2 domains)	0.428 (0.440)	12449	0.425 (0.465)	386	8–2.5
2	Conjugate-gradient atomic position refinement	0.379 (0.423)	23112	0.415 (0.441)	723	5–2.0
3	Modeling	0.356 (0.409)	23112	0.365 (0.436)	723	5–2.0
4	Conjugate-gradient atomic position refinement	0.330 (0.373)	23112	0.356 (0.376)	723	5–2.0
5	SA MD + conjugate gradient	0.281 (0.343)	23112	0.342 (0.369)	723	5–2.0
6	Modeling	0.284 (0.345)	23112	0340 (0.365)	723	5–2.0
7	Conjugate-gradient atomic position refinement	0.281 (0.343)	23112	0.337 (0.361)	723	5–2.0
8	*B* factors	0.277 (0.336)	23112	0.333 (0.358)	723	5–2.0
9	Conjugate-gradient atomic position refinement	0.273 (0.335)	23112	0.332 (0.358)	723	5–2.0
10	Torsion-angle MD	0.274 (0.332)	23112	0.329 (0.360)	723	5–2.0
11	Modeling	0.278 (0.329)	28878	0.331 (0.355)	893	5–1.86
12	Addition of 222 waters	0.249 (0.265)	28878	0.293 (0.313)	893	5–1.86
13	Conjugate gradient atomic position refinement	0.241 (0.253)	28878	0.289 (0.299)	893	5–1.86
14	*B* factors	0.235 (0.248)	28878	0.287 (0.297)	893	5–1.86
15	Modeling; addition of substrates	0.222 (0.234)	28878	0.268 (0.276)	893	5–1.86
16	Addition of 54 waters	0.212 (0.223)	28878	0.247 (0.254)	893	5–1.86
17	Conjugate-gradient atomic position refinement	0.207 (0.220)	28878	0.243 (0.253)	893	5–1.86
18	*B* factors	0.202 (0.215)	28878	0.235 (0.246)	893	5–1.86
19	Modeling	0.207 (0.217)	28878	0.234 (0.241)	893	5–1.86
20	Addition of 22 waters and 1 nitrate ion	0.206 (0.219)	28878	0.233 (0.238)	893	5–1.86
21	Conjugate-gradient atomic position refinement	0.200 (0.211)	28878	0.229 (0.234)	893	5–1.86
22	*B* factors	0.196 (0.209)	28878	0.224 (0.230)	893	5–1.86

† 3% of reflections were omitted for cross-validation. Reflections are counted according to the stated resolution limits used for refinement. The *R* factors in parentheses include all reflections down to a 100 Å low-resolution limit.

(iii) Progress was monitored in real and reciprocal space. In real space, the electron density of the substrates was monitored. They were not included in refinement or phasing. R^{free} was used as the reciprocal-space monitor (Brünger, 1997*a*), with a random 3% subset being omitted from both map calculation and refinement.

2.3.2. High-resolution refinement against 1.86 Å resolution cryo-data. The rate of convergence of the 2.5 Å resolution refinement with the room-temperature diffraction data was slowing. With $R^{free} = 0.339$, and before refining water molecule positions and individual *B* factors, the room-temperature data was replaced with the 1.86 Å resolution cryo-data, which is more suitable for atomic refinement, but forfeiting the experimental phases and inter-crystal averaging. Refinement was now exclusively in reciprocal space with *X-PLOR* (Brünger *et al.*, 1997), using a mixture of conjugate-gradient and molecular-dynamics (MD) protocols. Remodeling and water picking were performed with $2F_o - F_c$ maps. Water picking was primarily with the automatic routine of *X-PLOR*, selecting viable sites with peaks >1.5σ. Although the density became clear earlier, substrate molecules were added when R^{free} fell below 0.29 (Table 5). Progress to a precise 1.86 Å structure with $R^{free} = 0.224$ ($R^{work} = 0.196$) and good stereochemistry (Table 6; Fig. 3) was straightforward.

2.4. Biochemical significance of the structural results

Clear high-resolution electron density for the substrate analogues has allowed a precise model of the transition-state complex to be built. This has led to a substantially revised mechanism for the phosphagen kinases without the expected histidine acid–base catalyst. The structure is of wider interest, because it offers a very rare glimpse of a bimolecular enzyme reaction visualized at high resolution, without substantial perturbations of the substrates or active site. The replacement of the trigonal phosphorus with a nitrogen blocks the reaction with minimal stereochemical perturbation. The structure supported Jencks' theory (Page & Jencks, 1971) that bimolecular reactions could be catalyzed through reduction of the entropic component of the activation barrier, by constraining the orientation and position of substrates. It also provided the first direct visualization of a general bimolecular reaction mechanism proposed by Koshland (Dafforn & Koshland, 1971), in which reactions would be catalyzed by the precise alignment of substrates so that their orbitals were steered towards the optimal reaction trajectory.

3. *Post facto* analysis of real-space refinement

The primary objective had been the structure determination of arginine kinase, not a comparison of refinement methods. A number of different techniques had contributed to progress, and the critical nature of real-space refinement had not been established beyond doubt. Serendipity or different weighting could have given one method an advantage. To characterize specifically the role of real-space refinement, the initial refinement was repeated in controlled trials, incorporating refinement techniques that had newly become available.

3.1. Methods

To ensure a fair starting point for each refinement, the large domain was rebuilt from the Mi_b-CK structure positioned

Table 6
Structural statistics.

Protein atoms	2816
Solvent atoms	298
Substrate atoms	45
Ion atoms†	4
Average B for protein (Å^2)	16.6
Average B for substrates (Å^2)	9.2
⟨Deviation from ideal geometry⟩‡	
Bond lengths (Å)	0.007
Bond angles (°)	1.30
Torsion angles (°)	13.9
Residues in core ψ, φ regions§ (%)	91.7
Residues in disallowed regions (%)	0.0

† Some of water molecules picked by *X-PLOR* water pick turned out to be non-substrate nitrates bound to the protein. ‡ Compared with the Engh and Huber database (Engh & Huber, 1991). § Evaluated with *PROCHECK* (Laskowski *et al.*, 1993).

according to the molecular-replacement solution and rigid-body reciprocal-space refinement. In contrast to the actual structure determination, local (real-space) refinement was not used during rebuilding. This starting model was similar to the initial model in the actual structure determination with $R^{\rm free}$ = 0.539 and r.m.s. deviations from the final model of 2.1 Å (all atoms) and 1.2 Å (backbone).

All commonly used refinement methods were tested. Reciprocal-space methods included least-squares and maximum-likelihood optimization, gradient descent and/or simulated annealing, Cartesian or torsion-angle parameterization and used the programs *TNT* (Tronrud *et al.*, 1987) or *X-PLOR* (Brünger, 1992*b*). For each batch of refinement, the overall weight between stereochemistry and experimental fit was optimized with respect to decrease in $R^{\rm free}$ (Brünger, 1992*a*) and the weights between stereochemical terms were not adjusted. The real-space refinement is implemented as an addition to the *TNT* package (Chapman, 1995), using the same stereochemical restraining and optimizer programs as would be used in reciprocal space. Thus, some of the reciprocal-space refinements (including those in Table 7) were executed with *exactly* the same relative weights on different stereochemical terms as in real space.

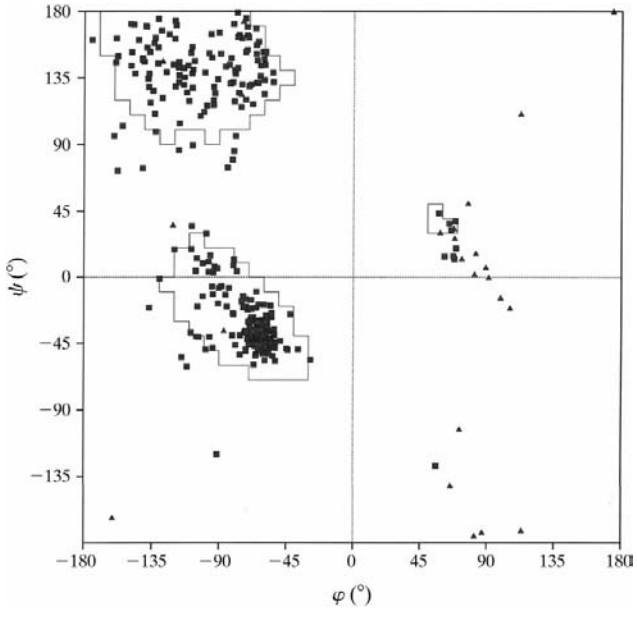

Figure 3
Ramachandran plot showing the agreement of observed backbone torsion angles with those which are stereochemically most favored (contour). Glycines are shown as triangles and other residues as squares. The figure was produced with *PROCHECK* (Laskowski *et al.*, 1993).

For the real-space refinement, the program *RSRef* was used (Chapman, 1995). This program calculates the electron-density map of a structure (or fragment) viewed within user-specified resolution limits. The electron-density contributions for each atom are calculated from the analytically evaluated Fourier transform of a resolution-truncated atomic scattering factor. The program scales the calculated density to the experimentally derived electron density and calculates the residual discrepancy and its first and second derivatives with respect to changes in each of the atomic coordinates or B factors. *RSRef* then passes this data to one of the crystallographic refinement programs, replacing the parts which normally analyze the residuals between structure and diffraction amplitudes, so that the fit of model to map and stereochemistry can be optimized simultaneously. *RSRef* has been implemented as a module for both *TNT* (Chapman, 1995; Tronrud *et al.*, 1987) and *X-PLOR* (Brünger, 1992*b*; Chen *et al.*, 1999*b*) to take advantage of popular gradient-descent or molecular-dynamics optimization methods. For the low-quality maps available for this work, molecular dynamics is not an advantage (Chen *et al.*, 1999*b*), so conjugate-gradient optimization was used with the *TNT*-based implementation.

During the actual structure determination, the model for the small domain was detrimental to atomic refinement by any method (see above). The difficulty of precisely orienting a small domain in such difficult structure determinations results, in part, from the overall large discrepancy between model and diffraction data. With improved consistency between large domain and diffraction data, improved placement of both domains might be possible. Thus, as in the actual structure determination, following tests of the atomic refinement of the large domain alone, the molecular-replacement polyglycine model for the small domain was added. The domains were refined in reciprocal space as two rigid bodies, and further atomic refinement was then attempted to test whether further progress was possible with a more complete model using either real or reciprocal-space refinement.

A common limitation of real-space refinement is the experimental phase and map quality (Chapman & Blanc, 1997; Chen *et al.*, 1999*b*). It is important to use the best unbiased map. For the *post facto* analysis, we used the map which was available at the corresponding stage of the actual structure determination. Phases had been calculated from the large domain molecular-replacement solution, (weak) isomorphous derivatives and refined by solvent flattening and inter-crystal averaging (see above). As discussed in §4, much is often gained by repeating real-space refinement after recalculating phases from a refined model. The objective of the *post facto* analysis was to test whether real-space refinement helped when the model was at its worst and other methods failed, not

Table 7
Post facto comparison of refinement strategies for the initial partial model.

In reciprocal space, the Cartesian conjugate-gradient method shown was the most successful of many attempts which included simulated-annealing and torsion-angle refinements.

Step	R^{free}	R^{work}
Starting model: large domain of Mi_b-CK structure (Fritz-Wolf *et al.*, 1996) placed in the arginine kinase form 2 unit cell according to the molecular-replacement solution.	0.543	0.522
Remodeling of large domain without refinement. (Small domain could not be modeled.) Product is the starting point of alternative refinement strategies.	0.539	0.518
Real-space refinement strategy		
Real-space refinement of large domain	0.539→0.519	0.481
Addition of Mi_bCK polyglycine model for small domain according to the molecular-replacement solution	0.519→0.506	0.484
Rigid-body refinement of the two domains in reciprocal space	0.506→0.486	0.480
Real-space refinement	0.486→0.478	0.475
Reciprocal-space refinement strategy using Cartesian conjugate-gradient methods		
Reciprocal-space refinement of large domain	0.539→0.530	0.447
Addition of polyglycine model for small domain (as above)	0.530→0.510	0.518
Rigid-body refinement	0.510→0.496	0.513
Further reciprocal-space refinement	All attempts diverge	

to conduct an exhaustive search for an efficient refinement protocol. No attempt was made to iteratively improve the map, and the results are, therefore, a conservative illustration of what might be possible.

3.2. Results and discussion

The statistics for both the real- and reciprocal-space refinements of the *post facto* analysis are slightly better than those of the actual structure determination (Table 4). This is a result of more careful optimization of the refinement weights than in the actual structure determination. With a poor starting model, optimization against R^{free} (Brünger, 1992*a*) indicated that it was sometimes advantageous to allow more than the usual deviation from ideal stereochemistry.

Attempts at refining the large domain in reciprocal space led to more modest improvement than in real space. R^{free} dropped to 0.531 by conjugate-gradient methods (compared to 0.518 in real space), 0.535 by Cartesian maximum-likelihood (ML) methods and increased to 0.561 by torsion-angle ML methods (Brünger *et al.*, 1997, 1998; Pannu & Read, 1996). In all reciprocal refinement tests, there was evidence of substantial overfitting with working R factors falling to 0.468, 0.470 and 0.451, respectively (0.063–0.11 below R^{free}; Brünger, 1997*a*), whereas after real-space refinement R^{work} and R^{free} differed by only 0.005.

The controlled *post facto* comparisons confirmed impressions gained during the actual structure determination. Real-space refinement of the large domain gave modest improvement (Table 7) which allowed further rigid-body improvement. (The molecular-replacement solution had previously been rigid-body refined.) The real-space and rigid-body refinements led to a total improvement in the R^{free} of 0.061 (to 0.478), with drops of 0.028, 0.013 and 0.20 attributable to atomic refinement, addition of the small domain and rigid-body refinement, respectively. Comparisons with the final structure showed a 10% reduction in r.m.s. coordinate error. Conventional conjugate-gradient refinement (which had been the most successful of the reciprocal-space methods in the actual structure determination) led to a smaller 5% improvement in coordinate error with R^{free} dropping 0.043, of which 0.02 and 0.014 were attributable to addition of the small domain and rigid-body refinement and only 0.009 was attributable to atomic refinement. Furthermore, with the reciprocal-space refinement, there was substantial overfitting which could exacerbate model bias. By contrast, with the real-space strategy, the close correspondence of R^{free} and R^{work} confirms that there is essentially no overfitting, making it a more suitable phasing model for further progress.

4. Conclusions

This structure determination was typical of a large subset in which marginal phasing leads to sometimes intractable challenges in building an atomic model which will converge upon the correct structure during atomic refinement. As is often the case, the molecular replacement was made more difficult by the suitability of a probe molecule with substantially different conformation. Experimental phase information was important in attaining sufficient confidence in ambiguous molecular-replacement solutions. Use of all available phasing information, including inter-crystal-form averaging, was important in obtaining an interpretable map and in starting the refinement.

The structure determination was made difficult because of the low-quality phases. It may then seem counter-intuitive that it is advantageous to use a refinement method which is implicitly dependent upon the phases. The phases do provide additional (implicit) restraints which increase the effective data-to-parameter ratio and may help to stabilize convergence. However, it is more likely that the value of the real-space method is in the localization of the refinement. The refinement of each atom is dependent only on the electron density and neighboring atoms in the immediate vicinity. The only way that remote parts of the structure may affect refinement is through the quality of phases that might be calculated from all atoms. Even then, the interactions are tenuous when phase recalculation is infrequent, uses 'omit' techniques (Bhat & Cohen, 1984) and incorporates model-independent information. This is in sharp contrast to reciprocal-space methods, in which the interaction between

remote parts of the structure is strong and immediate (Hodel *et al.*, 1992). In reciprocal space, atoms are optimized to give the best overall agreement with the structure amplitudes that are a function of all atoms. Thus, if a significant part of the structure is absent or seriously in error (as in this case), other atoms will be likely to refine to (slightly) incorrect positions to compensate for the erroneous region and improve the overall agreement with the diffraction amplitudes. Maximum-likelihood methods (Murshudov *et al.*, 1997; Pannu & Read, 1996), which continue to be developed, may approximate the 'average' effects of missing or erroneous model. They should, in principle, be an improvement upon conventional atomic refinement. However, for the most extreme cases, use of a strictly *local* real-space refinement, which avoids long-range interactions, is likely to remain superior to probabilistic mitigation of the effects of model deficiencies upon structure factors.

The effect of real-space refinement on the coordinates of a crude model can be slight. For the actual structure determination, it is not clear how much coordinate improvement there would have been had the steps of modeling and refinement been kept distinct, but it would be likely to have been small. Improvements of the fit of model to density are only just noticeable (Fig. 1). For the *post facto* analysis, the decrease in coordinate error was 10%. Given then that coordinate errors can only be evaluated in retrospect, how is one to judge whether real-space refinement is helping?

For both the real-life and *post facto* refinements, $R^{\rm free}$ (Brünger, 1997*a*) was the most robust indicator of improvement. The potential advantages of initial real-space refinement can be quickly ascertained through a short series of low-resolution trial refinements in which real-space refinement is compared with the reciprocal-space method of choice. First, the appropriate weighting between diffraction and stereochemical terms needs to be established approximately for each method of refinement. This should be performed by searching for lowest $R^{\rm free}$ following five or six cycles of refinement (Brünger, 1992*a*), as the best weights may differ from those appropriate for better models. Three protocols should then be tested: (*a*) real-space refinement alone, (*b*) reciprocal-space refinement alone and (*c*) real-space followed by reciprocal-space refinement, as in both this and prior work initial real-space refinement has been found to improve the convergence of subsequent reciprocal-space refinement (Chapman & Blanc, 1997). For the real-space refinement, the highest quality *unbiased* map should be used. If calculated phases are to be used, the model should not have previously been refined against the current amplitudes, and omit techniques should be used (Bhat & Cohen, 1984). Eight to ten cycles of refinement should be sufficient to reach convergence. The outcomes of the three protocols can be judged according to the decrease in $R^{\rm free}$, a measure of model quality, and the lack of decrease in the difference between $R^{\rm free}$ and $R^{\rm work}$, an indication of overfitting and potential future phase bias (Brünger, 1997*a*). While free R factors calculated by standard methods (Brünger, 1997*a*) were sufficient in the AK structure determination, it has been noted that omission of test reflections from map calculation has a detrimental effect upon real-space refinement and also artifactually increases $R^{\rm free}$ (Chen *et al.*, 1999*a*). Improved methods of calculating $R^{\rm free}$ mitigate these effects while retaining strict separation between test and working data, and offer a fairer appraisal of the potential benefits of real-space refinement (Chen *et al.*, 1999*a*). In summary, a few hours of test refinements can be used to confirm the benefits of real-space refinement in terms of model improvement without overfitting, as judged by calculation of free and working R factors.

The *post facto* analyses focused on the very first step of refinement, confirming that real-space methods can be the most appropriate in particularly challenging cases. A batch of real-space refinement is often limited by map quality, but no attempt was made to fully exploit the technique by alternating refinement with map recalculation using phases calculated from the refined model. It has previously been demonstrated that an effective way of extending the power of automatic refinement is to repeat a few times a cycle of real-space refinement to convergence, reciprocal-space refinement and then map recalculation (Chapman & Blanc, 1997). The real-space refinement has a wide convergence radius and less propensity for overfitting in the susceptible early stages. Reciprocal-space refinement frees the model from the limitations of the phases. Subsequent real-space refinement mimics in part what would usually be accomplished with labor-intensive interactive modeling. However, it was noted in the previous work that once maps with calculated phases are used, overfitting can become a problem with real-space refinement as it is in reciprocal-space refinement (Chapman & Blanc, 1997) as models are adjusted to fit phase-biased maps. In the prior work, little attempt was made to reduce phase bias: $2F_o - F_c$ maps were used. In the actual AK structure determination reported here, greater efforts were made to reduce phase bias. Phases were calculated by back-transformation of $\sigma_{\rm A}$-weighted omit maps (Bhat & Cohen, 1984; Read, 1985) combined with isomorphous replacement data, and were further refined away from model biased values by density modification and inter-crystal averaging. The result was that real-space methods continued to be valuable through the intermediate stages of refinement, and that evidence of overfitting ($R^{\rm free} - R^{\rm work}$) was less than typical (Table 4). Thus, the *post facto* studies offer a conservative impression of what might be possible. The best results are likely to come by cycling through real- and reciprocal- space refinements and map recalculation, while taking precautions to minimize phase bias.

We further emphasize that application of real-space methods does not exclude subsequent reciprocal-space refinement. In fact, real-space refinement often has a positive impact on subsequent reciprocal-space refinement and, in many trials in the authors' laboratory, has never been found to be detrimental. Thus, the methods used in this structure determination should be considered those of first resort in cases where it is possible to build only a poor or partial atomic model into an initial electron-density map.

The structure determination of arginine kinase was supported by the American Heart Association, Florida Division and the National Institutes of Health (GM 55837). Methodological work was supported by the National Science Foundation (BIR94-18741 & DBI98-08098). Refinement programs are available from http://www.sb.fsu.edu/~rsref.

References

Bhat, T. N. & Cohen, G. H. (1984). *J. Appl. Cryst.* **17**, 244–248.

Blanc, E., Chen, Z. & Chapman, M. S. (1998). *Direct Methods for Solving Macromolecular Structures*, edited by S. Fortier, pp. 513–519. Dortrecht: Kluwer.

Brünger, A. T. (1992*a*). *Nature (London)*, **355**, 472–475.

Brünger, A. T. (1992*b*). *X-PLOR Version* 3.1. *A System for X-ray Crystallography and NMR*. Yale University Press, New Haven, Connecticut, USA.

Brünger, A. T. (1997*a*). *Methods Enzymol.* **276**, 558–580.

Brünger, A. T. (1997*b*). *Methods Enzymol.* **277**, 366–396.

Brünger, A. T., Adams, P. D., Clore, G. M., Gros, P., Gross-Kunstleve, R. W., Jiang, J.-S., Kurzewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst.* D**54**, 905–921.

Brünger, A. T., Adams, P. D. & Rice, L. M. (1997). *NATO Advanced Study Institute on Direct Methods for Solving Macromolecular Structures*, pp. 149–163. Dordrecht: Kluwer.

Brünger, A. T., Kuriyan, J. & Karplus, M. (1987). *Science*, **235**, 458–60.

Carter, C. W. Jr, Ilyin, V., Huang, X. & Li, G. (1996). *Acta Cryst.* A**52**, C82.

Chapman, M. S. (1995). *Acta Cryst.* A**51**, 69–80.

Chapman, M. S. & Blanc, E. (1997). *Acta Cryst.* D**53**, 203–206.

Chen, Z., Blanc, E. & Chapman, M. S. (1999*a*). *Acta Cryst.* D**55**, 219–224.

Chen, Z., Blanc, E. & Chapman, M. S. (1999*b*). *Acta Cryst.* D**55**, 464–468.

Cowtan, K. D. & Main, P. (1993). *Acta Cryst.* D**49**, 148–157.

Cowtan, K. D. & Main, P. (1996). *Acta Cryst.* D**52**, 43–48.

Dafforn, A. & Koshland, D. E. Jr (1971). *Proc. Natl Acad. Sci. USA*, **68**, 2463–2467.

Diamond, R. (1971). *Acta Cryst.* A**27**, 436–452.

Engh, R. A. & Huber, R. (1991). *Acta Cryst.* A**47**, 392–400.

Forstner, M., Kriechbaum, M., Laggner, M. P. & Wallimann, T. (1996). *J. Mol. Struct.* **383**, 217–227.

Fritz-Wolf, K., Schnyder, T., Wallimann, T. & Kabsch, W. (1996). *Nature (London)*, **381**, 341–345.

Hendrickson, W. A. & Ogata, C. M. (1997). *Methods Enzymol.* **276**, 494–523.

Hodel, A., Kim, S.-H. & Brünger, A. T. (1992). *Acta Cryst.* A**48**, 851–858.

Jones, T. A. & Kjeldgaard, M. (1997). *Methods Enzymol.* **277**, 173–208.

Jones, T. A. & Liljas, L. (1984). *Acta Cryst.* A**40**, 50–57.

Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* A**47**, 110–119.

Kenyon, G. L. & Reed, G. H. (1983). *Adv. Enzymol.* **54**, 367–426.

Kleywegt, G. J. & Jones, T. A. (1993). *ESF/CCP4 Newslett.* **28**, 56–59.

Knight, S. (1989). PhD thesis. Swedish University of Agricultural Sciences, Uppsala, Sweden.

Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.

McPherson, A. (1973). *J. Mol. Biol.* **82**, 83–86.

Murshudov, G., Vagin, A. & Dodson, E. (1997). *Acta Cryst.* D**53**, 240–255.

Navaza, J. (1994). *Acta Cryst.* A**50**, 157–163.

Navaza, J. & Saludjian, P. (1997). *Methods Enzymol.* **277**, 581–594.

Otwinowski, Z. (1991). *Proceedings of the CCP4 Study Weekend*, edited by W. Wolf, P. R. Evans & A. G. W. Leslie, pp. 80–86. Warrington: Daresbury Laboratory.

Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.

Page, M. I. & Jencks, W. P. (1971). *Proc. Natl Acad. Sci. USA*, **68**, 1678–1683.

Pannu, N. S. & Read, R. J. (1996). *Acta Cryst.* A**52**, 659–668.

Ramachandran, G. N. & Srinivasan, R. (1970). In *Fourier Methods in Crystallography*, edited by M. J. Buerger. New York: John Wiley.

Read, R. J. (1985). *Acta Cryst.* A**42**, 140–149.

Rice, L. M. & Brünger, A. T. (1994). *Proteins Struct. Funct. Genet.* **19**, 277–290.

Rodgers, D. W. (1994). *Structure*, **2**(12), 1135–1140.

Rossmann, M. G. (1972). *The Molecular Replacement Method.* New York: Gordon and Breach.

Stout, G. H. & Jensen, L. H. (1989). *X-ray Structure Determination – A Practical Guide*, 2nd ed. New York: Macmillan.

Strong, S. J. & Ellington, W. R. (1996). *Comput. Biochem. Physiol.* **113**B, 809–816.

Terwilliger, T. C., Kim, S.-H. & Eisenberg, D. S. (1987). *Acta Cryst.* A**43**, 1–5.

Tronrud, D. E., Ten Eyck, L. F. & Matthews, B. W. (1987). *Acta Cryst.* A**43**, 489–501.

Wang, B.-C. (1985). *Methods Enzymol.* **115**, 90–112.

Zhang, X.-J. & Matthews, B. W. (1994). *Acta Cryst.* D**50**, 675–686.

Zhou, G., Parthasarathy, G., Somasundaram, T., Ables, A., Roy, L., Strong, S. J., Ellington, W. R. & Chapman, M. S. (1997). *Protein Sci.* **6**, 444–449.

Zhou, G., Somasundaram, T., Blanc, E., Parthsarathy, G., Ellington, W. R. & Chapman, M. S. (1998). *Proc. Natl Acad. Sci. USA*, **95**, 8449–8454.