Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

A subgroup algorithm to identify cross-rotation peaks consistent with non-crystallographic symmetry

Ryan H. Lilien,[a,b,c] Chris Bailey-Kellogg,[d] Amy C. Anderson[c,e]* and Bruce R. Donald[a,c,e,f]*

[a]Dartmouth Computer Science Department, Hanover, NH 03755, USA, [b]Dartmouth Medical School, Hanover, NH 03755, USA, [c]Dartmouth Center for Structural Biology and Computational Chemistry, Hanover, NH 03755, USA, [d]Purdue Department of Computer Sciences, West Lafayette, IN 47907,USA, [e]Dartmouth Department of Chemistry, Hanover, NH 03755, USA, and [f]Dartmouth Department of Biological Sciences, Hanover, NH 03755, USA

Correspondence e-mail: amy.c.anderson@dartmouth.edu, bruce.r.donald@dartmouth.edu

Received 5 December 2003
Accepted 24 March 2004

Molecular replacement (MR) often plays a prominent role in determining initial phase angles for structure determination by X-ray crystallography. In this paper, an efficient quaternion-based algorithm is presented for analyzing peaks from a cross-rotation function in order to identify model orientations consistent with proper non-crystallographic symmetry (NCS) and to generate proper NCS-consistent orientations missing from the list of cross-rotation peaks. The algorithm, *CRANS*, analyzes the rotation differences between each pair of cross-rotation peaks to identify finite subgroups. Sets of rotation differences satisfying the subgroup axioms correspond to orientations compatible with the correct proper NCS. The *CRANS* algorithm was first tested using cross-rotation peaks computed from structure-factor data for three test systems and was then used to assist in the *de novo* structure determination of dihydrofolate reductase–thymidylate synthase (DHFR-TS) from *Cryptosporidium hominis*. In every case, the *CRANS* algorithm runs in seconds to identify orientations consistent with the observed proper NCS and to generate missing orientations not present in the cross-rotation peak list. The *CRANS* algorithm has application in every molecular-replacement phasing effort with proper NCS.

1. Introduction[1]

When the structure of an homologous protein is known, initial phases for the diffraction data can often be determined using the technique of molecular replacement (Rossmann & Blow, 1962; Crowther & Blow, 1967; Rossmann, 1990). To use initial phases from an homologous model, each copy of the molecular-replacement model (henceforth called the *model* for brevity) must be properly oriented and translated within the asymmetric unit (Tong & Rossmann, 1990; Rossmann & Blow, 1962; Crowther & Blow, 1967; Rossmann, 1990). The task of properly orienting and translating each model is facilitated by exploiting the additional constraint provided by proper non-crystallographic symmetry (NCS).

Traditionally, the initial presence and degree (*i.e.* threefold, fourfold *etc*.) of NCS is first identified by a self-rotation function (Rossmann & Blow, 1962), while the orientations of each copy of the model are subsequently identified using the cross-rotation function (Rossmann & Blow, 1962). Ideally, after *n*-fold NCS has been identified, the peaks of the cross-rotation function will possess two desirable properties. First,

[1] Abbreviations used: *SO*(3), group of three-dimensional rotations; MR, molecular replacement; model, molecular-replacement model; DHFR, dihydrofolate reductase; TS, thymidylate synthase; ChDHFR-TS, *Cryptosporidium hominis* DHFR-TS; LmDHFR-TS, *Leishmania major* DHFR-TS; PcTS, *Pneumocystis carinii* TS; NCS, non-crystallographic symmetry (proper, unless otherwise noted); RMSD, root-mean-square distance.

the top n cross-rotation peaks should have rotation-function scores significantly higher than the rest and, second, these top peaks should correspond to the correct NCS-consistent orientations. A complication arises when the model is partial (Oh, 1995), when the model shares only moderate structural similarity with the crystallized molecule or when the degree of NCS is high and each model represents only a small fraction of the unit cell's molecular mass. Empirically, in these situations the n top-scoring rotations from a cross-rotation function search may not correspond to the n correct (NCS-consistent) rotations. In fact, some of the correct rotations may not even appear in the cross-rotation peak list, further complicating the search for the correct model orientations. For complex systems it is therefore possible that the wrong cross-rotation peaks are selected for use with the computationally expensive translation function. We have developed an algorithm to (i) compute which cross-rotation peaks generate proper NCS-consistent model orientations and (ii) generate proper NCS-consistent model orientations not specified by rotations in the cross-rotation peak list.

The *CRANS* (cross-rotation analysis) algorithm identifies and computes sets of rotations (set R in Fig. 1) that, when applied to the model, produce orientations consistent with the NCS. We call these sets of rotations *NCS-consistent rotation sets*. *CRANS* therefore reduces the time required to obtain initial phases in two ways. First, by correctly identifying NCS-consistent rotation sets among the rotations identified by the cross-rotation function, we reduce the number of improperly oriented models and thereby the total number of translation searches required to generate initial phases. Second, in the case where one or more NCS-consistent rotations are absent from the cross-rotation peak list, *CRANS* (i) informs the crystallographer that a peak is missing without requiring them to perform translation searches on each cross-rotation peak and (ii) computes the missing NCS-consistent rotations.

The core of the *CRANS* algorithm analyzes the set of *rotation differences* to identify finite sets of rotations that satisfy a point group with a single rotation axis. Given a set of w cross-rotation peaks (rotations) $C = \{r_1, r_2, \ldots, r_w\}$, where r_i is an element of the group of three-dimensional rotations, *CRANS* examines the w^2 rotation differences d_{ij}, where $d_{ij} = r_i^{-1} r_j$ and r_i^{-1} is the inverse of r_i (such that $r_i^{-1} r_i$ is the identity). Conceptually, the rotation difference d_{ij} is the rotation that rotates the model oriented by r_i into the model oriented by r_j (Fig. 1). We can test the NCS-consistency of a set of cross-rotation peaks by examining their rotation differences and verifying that they share a common axis and form a finite subgroup (Appendix *A*). In the event of missing rotations, the *CRANS* algorithm completes each partial set of identified NCS-consistent rotations by generating missing rotations with quaternions (Hamilton, 1969; Salamin, 1979). Although rotation differences can be computed using a number of rotation representations (Evans, 2001), we chose to use quaternions because they have a single compact representation with explicit axis and angle components, are easily composed, are free of singularities, and represent a uniform parameterization of rotation space.

When the model P is a homodimer, the search for NCS-consistent rotation sets is more complex because the orientation rP (where rP is the result of rotating protein P by rotation r) is equivalent to the orientation rfP, where f is the 180° rotation around the dimer twofold axis (Fig. 2). This rotational degeneracy increases the difficulty of the search because the rotational relationships between the cross-

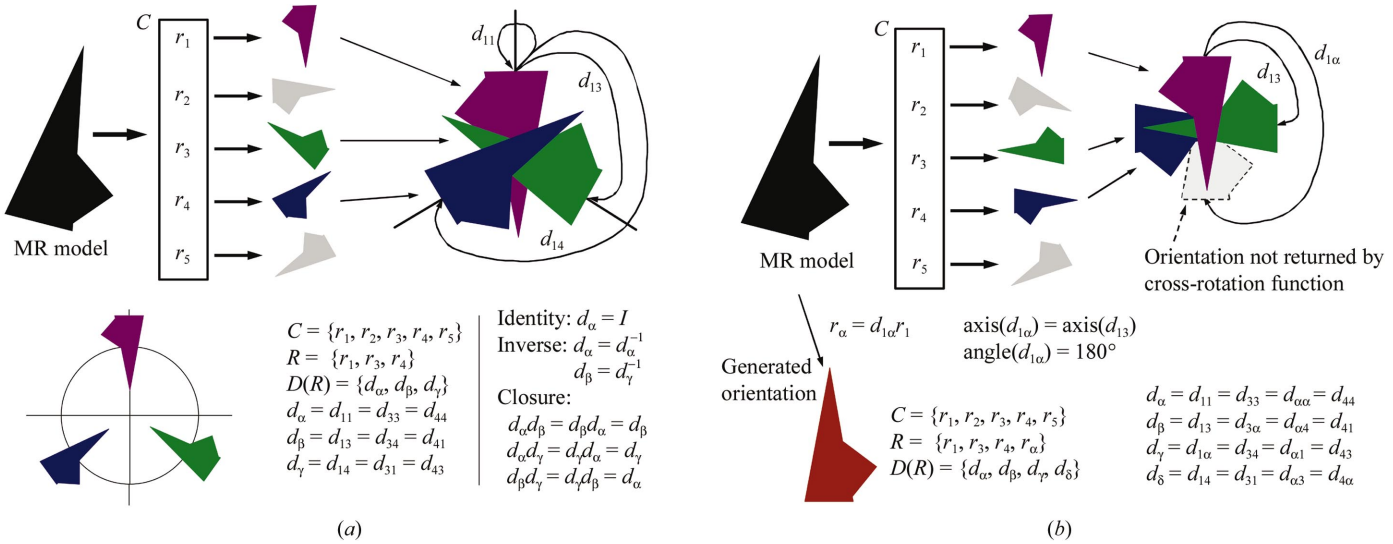

Figure 1
Two-dimensional examples for threefold (*a*) and fourfold (*b*) NCS. (*a*) A model is shown with the results of a simplified cross-rotation search C containing only five rotations. Orientations corresponding to rotations r_1 (purple), r_3 (green) and r_4 (blue) form an NCS-consistent rotation set R. For clarity, only a few rotation differences (d_{11}, d_{13} and d_{14}) are shown in the upper right overlapping orientation figure. The rotation differences $D(R)$ form a complete rotation difference set and satisfy the group properties of associativity (not shown), identity, inverse and closure. (*b*) A fourfold NCS example is shown using similar notation to (*a*). In this example, only three of the four NCS-consistent rotations (r_1, r_3 and r_4) are contained in the cross-rotation peak list C. The missing rotation $r_\alpha = d_{1\alpha} r_1$ is computed using $d_{1\alpha}$ defined by the axis of the three identified NCS-consistent rotations and the missing angle. The now complete NCS-consistent rotation set R has a complete rotation difference set $D(R)$ which satisfies the subgroup properties.

rotation peaks must be examined modulo *f*. Thus, the dimer axis is explicitly considered by *CRANS* when using a homodimer model.

Although the *CRANS* algorithm only requires a cross-rotation peak list as input, if the NCS axis and/or degree are known (*i.e.* from the self-rotation search) they can be used as an additional constraint in identifying rotation subgroups. In addition to searching for NCS-consistent rotation sets satisfying one particular symmetry, the *CRANS* algorithm may also be run in 'scan' mode, where the algorithm consecutively attempts to identify subgroups of rotation differences consistent with a user-specified range of NCS degree (*i.e.* threefold, fourfold, ..., *n*-fold).

Previous work in automated cross-rotation peak analysis includes the programs *RFCORR* and *MOLREP* (Vagin & Teplyakov, 1997; both distributed as part of *CCP*4; Collaborative Computational Project, Number 4, 1994). *RFCORR* and *MOLREP* both compute the set of rotation differences but leave the task of identifying sets of NCS-consistent rotations to the user; *CRANS* automates this process and identifies NCS-consistent rotation sets even when there are missing cross-rotation peaks.

In summary, the *CRANS* algorithm analyzes the output of a cross-rotation search and makes the following contributions:

(i) *CRANS* computes all sets of *n*-fold NCS-consistent rotation sets, even in the presence of potentially missing peaks. The missing peaks (*i.e.* those not found by the cross-rotation function) required to form a *complete* (*i.e.* no missing rotations) NCS-consistent rotation set are also computed by *CRANS*. The computed missing peaks can be used to orient the model prior to performing a translation search.

(ii) During a traditional manual examination of cross-rotation peaks, one typically only examines the top *p* cross-rotation peaks. In contrast, the *CRANS* algorithm performs an exhaustive search over all peaks in the cross-rotation function, checking each rotation for inclusion into an NCS-consistent rotation set.

(iii) *CRANS* computes the NCS axis for each identified NCS-consistent rotation set. If the NCS axis is known from the self-rotation map, *CRANS* can limit its search to return only those rotation sets with NCS axes close to the known NCS axis.

(iv) *CRANS* can also be run in 'scan' mode, where the algorithm sequentially searches a range of degrees of NCS and attempts to identify NCS-consistent rotation sets for each degree of symmetry.

(v) Finally, *CRANS* can perform the four above listed tasks when the model is a homodimer. In our experience, the ability to handle homodimer models was vital in solving the structure of dihydrofolate reductase–thymidylate synthase (DHFR-TS) from *Cryptosporidium hominis* (O'Neil *et al.*, 2003; §3.4).

The paper is organized as follows. In §2 we describe the *CRANS* algorithm, present the algorithm's runtime complexity and describe the data and pre-processing used for each test system. In §3 and §4 the performance of *CRANS* is demonstrated by analyzing six cross-rotation peak lists generated from six models against four protein crystal systems. Firstly, we describe the performance of the *CRANS* algorithm on three test systems of solved crystal structures containing threefold (2-keto-3-deoxy-6-phosphogluconate aldolase; PDB code 1fq0; Wymer *et al.*, 2001), fivefold (cholera toxin B subunit; PDB code 1chp; Merritt *et al.*, 1995) and sevenfold (Gp31 co-chaperonin; PDB code 1g31; Hunt *et al.*, 1997) NCS. Then, in §3.4 we describe how we used *CRANS* to solve the *de novo* structure of dihydrofolate reductase–thymidylate synthase (DHFR-TS) from *C. hominis* (ChDHFR-TS; PDB code 1qzf; O'Neil *et al.*, 2003), a homodimer, where we used *CRANS* to identify the cross-rotation peaks consistent with fivefold NCS. Despite the fact that three of the six tested cross-rotation peak lists were missing at least one and in one instance as many as three NCS-consistent rotations, in all cases *CRANS* was able to correctly identify NCS-consistent rotations and generate all missing rotations. These results support the general applicability of the *CRANS* algorithm for use in analyzing cross-rotation search results for systems with NCS. In §5 we discuss the use of *CRANS* in crystals with improper NCS, multiple point-group symmetries and parallel crystallographic and non-crystallographic axes.

2. Methods

2.1. *CRANS* algorithm

The *CRANS* algorithm takes a list of cross-rotation peaks as input and identifies sets of cross-rotation peaks that generate orientations of the model related by *n*-fold NCS (Appendix *B*). In describing the *CRANS* algorithm, we define the func-

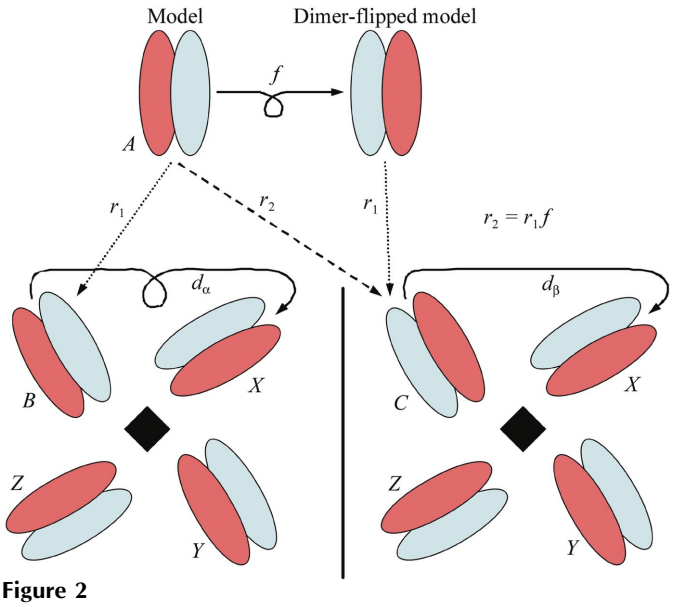

Figure 2
Homodimer-based rotations. The model in orientation *A* can be rotated by either r_1 or r_2 to assume orientations *B* or *C*. While both orientations *B* and *C* can form a fourfold NCS (modulo the dimer flip) with orientations *X*, *Y* and *Z*, the rotations d_α and d_β are not equal. The rotation difference d_α is not consistent (*i.e.* does not form a subgroup) with the partial rotation difference set generated from orientations *X*, *Y* and *Z*. Thus, it becomes important that when the model is a homodimer, for every rotation r_1 we generate the dimer-flip-related rotation $r_2 = r_1 f$ (where the rotation *f* flips the model along its twofold dimer axis).

tions axis(·) and angle(·) to return the axis and angle, respectively, of a given rotation. We define the n angles specified by $360t/n$ ($t \in \{0, 1, \ldots, n-1\}$) as *symmetry angles*. Let m be the maximum number of missing rotations allowed by the user.

The *CRANS* algorithm is divided into four stages: process input, filter, partition and patch. In the process-input step, all w *original rotations* (rotations from the cross-rotation function) $C = \{r_1, r_2, \ldots, r_w\}$ are read from an input file and converted to quaternions. The *CRANS* input format is described in Appendix *B*. After reading the cross-rotation peak-list input file, *CRANS* computes all w^2 rotation differences, d_{ij}.

In the filter stage, all rotation differences with angle(d_{ij}) differing from a symmetry angle by more than the user-defined angular threshold τ_{angle} are discarded. The intuition behind this filtering step is that two rotations r_i and r_j that are both members of the correct NCS-consistent rotation set should have a rotation difference angle(d_{ij}) approximately equal to a symmetry angle.

The core of the algorithm occurs in the partition stage. In general, if rotation differences d_{ij} and d_{kl} are both members of the same NCS-consistent rotation set, then axis(d_{ij}) $\approx$ axis(d_{kl}). Therefore, in this stage, each remaining (*i.e.* unfiltered by step 2) rotation difference d_{ij} is clustered with all rotation differences d_{ik} that share a common rotation axis. That is, the differences d_{ij} and d_{ik} are assigned to the same set if $|\theta_{ij,ik}| \leq \tau_{\text{axis}}$, where $\theta_{ij,ik}$ ($0 \leq \theta_{ij,ik} \leq 90^\circ$) is the angle between axis(d_{ij}) and axis(d_{ik}) and τ_{axis} is the user-defined axis similarity threshold. Sets of rotation differences with more than m missing *distinct rotations* are eliminated. In every set S_i of rotation differences of the form d_{ij} (where i is fixed and j varies, *i.e.* all rotation differences involve rotation r_i) the number of distinct rotations equals the number of unique orientations generated by applying $d_{ij} \in S_i$ to r_iP, where P is the model. Because multiple rotations in a single rotation difference set might generate the same orientation, all possible subsets of rotation differences containing at most m missing distinct rotations are computed. At this point, each set S_i of rotation differences has generated zero or more partial rotation difference sets Δ'. For each partial rotation difference set Δ', let $D^{-1}(\Delta')$ be the set of all original cross-rotation peaks used to generate rotations in Δ' and let $\mathbf{v}$ be the average axis of all rotation differences formed by pairs of rotations in $D^{-1}(\Delta')$. The consistency of the NCS axis for each rotation difference is checked by comparing the axis of rotation difference d_{ab} to the average axis $\mathbf{v}$ for all r_a, $r_b \in D^{-1}(\Delta')$ to ensure $|\theta_{ab,\mathbf{v}}| \leq \tau_{\text{axis}}$, where $\theta_{ab,\mathbf{v}}$ ($0 \leq \theta_{ab,\mathbf{v}} \leq 90^\circ$) is the angle between axis(d_{ab}) and the average axis $\mathbf{v}$. Rotation difference sets not passing this filter are eliminated. After computing all rotation sets, those sets that are subsets of other, more complete, remaining rotation sets are removed.

In the final stage of the algorithm, the patch stage, missing rotations for each remaining rotation set R are computed. For example (Fig. 1*b*), if $n = 4$, $m = 1$ and a rotation set R contains three rotations r_1, r_3 and r_4 such that axis(d_{13}) = axis(d_{14}) = axis(d_{34}) and angle(d_{13}) = 90°, angle(d_{14}) = 270°, angle(d_{34}) = 180°, then the missing rotation $r_\alpha = d_{1\alpha}r_1$ is computed, where $d_{1\alpha}$ is the quaternion representing a rotation of 180° (the missing angle in this example) around axis(d_{13}) (Fig. 1*b*). By construction, $D(R)$ will be a subgroup (see Appendix *A*) and thus R will be an NCS-consistent rotation set.

The NCS-consistent rotation sets identified by *CRANS* are sorted first on their completeness (*i.e.* sets with fewer missing rotations are listed first) and second on a quality score. Two quality scores are implemented in *CRANS*. In the first method, RF-Scoring, the score of each NCS-consistent rotation set is computed as the sum of the rotation-function scores (RF-Scores) of the included cross-rotation peaks. In the second method, axis–angle deviation scoring, the score s of each NCS-consistent rotation set R is computed as the average angular deviation between the axis of each rotation difference and $\mathbf{v}$ plus the average angular deviation between each rotation difference and its closest symmetry angle,

$$s = \frac{2}{n(n+1)} \sum_{i=1}^{n} \sum_{j=i+1}^{n} \left(\theta_{ij,\mathbf{v}} + \min_{t \in Z_n} |\text{angle}(d_{ij}) - 360t/n| \right), \quad (1)$$

where n is the number of rotations in R, $\theta_{ij,\mathbf{v}}$ is the angle between axis(d_{ij}) and the average axis $\mathbf{v}$ and $Z_n = \{0, 1, \ldots, n-1\}$.

When the model is a homodimer, one change is made to the process-input stage of the above algorithm. First, the dimer axis is computed from the model structure and is used to compute f, the 180° rotation around the twofold dimer axis (Fig. 2). The rotation f is then used to generate a dimer-flipped version of each original rotation, which is added to the list of rotations utilized by subsequent stages of the algorithm. Formally, the set of cross-rotation peaks C is replaced by $C \cup \{rf \mid r \in C\}$. The algorithm then proceeds as described above.

Because *CRANS* does not directly utilize knowledge of crystallographic symmetry, it unable to expand the cross-rotation peaks into their crystallographic equivalents. Therefore, NCS-consistent orientations generated by crystallographic symmetry must be identified during the cross-rotation search. The *CNS* (Brünger *et al.*, 1998) cross-rotation search used in all our examples performs this crystallographic expansion. Therefore, *CRANS* can directly analyze the cross-rotation search output.

2.2. Complexity

Reading, conversion to quaternions and computation of the inverse of each of the w cross-rotation peaks requires $O(w)$ time. Computing all rotation-pair differences requires $O(w^2)$ time. Filtering the rotation-pair differences (for those near the n symmetry angles) requires time $O(nw^2)$. Formation of rotation-difference sets sharing a common axis takes a worst-case time of $O(w^4)$, which occurs when all w^2 rotation differences have passed the previous symmetry-angle filter. In practice, the number of remaining differences is quite small and only a fraction of the original rotation differences are used to compute rotation-difference sets. Creation and patching of the final sets consistent with the desired NCS requires time

$O(g)$, where g is the number of NCS-consistent rotation sets. Therefore, the *CRANS* algorithm can search for NCS symmetry among w cross-rotation peaks with an expected runtime of $O(nw^2 + g)$ and a worst-case runtime of $O(w^4 + g)$. In practice, g is a small constant and we can reduce the runtime to an expected $O(nw^2)$ and a worst-case $O(w^4)$. Our implementation of *CRANS* requires only seconds to search for up to eightfold NCS on lists of 120 cross-rotation peaks using an Athlon-based processor.

2.3. Source of data and preprocessing

Structure factors for the three test systems 1fq0, 1chp and 1g31 were obtained from the Protein Data Bank (PDB; Berman *et al.*, 2000) and converted into *CNS* (Brünger *et al.*, 1998) format. The first chain from the crystal structure (representing a single monomer) was extracted and used as the model for molecular replacement. A cross-rotation search was performed with *CNS* using the following default parameters: resolution limits, 15–4 Å; data cutoff criteria, 0.0; r.m.s. outlier cutoff, 1000; bins for resolution-dependent operations, 10; atoms, known and not hydrogen; scoring function, fastdirect; use automatically determined asymmetric unit; fastdirect grid factor, 5; fastdirect maximum number of peaks, 20; cluster threshold, 10°; maximum number of coarse peaks to analyze in fine grid search, 20. The generated cross-rotation function peak list was then analyzed by *CRANS*. The same axis and angle tolerances $\tau_{\rm axis} = 4.5^\circ$ and $\tau_{\rm angle} = 5.0^\circ$ were used in all *CRANS* analyses. Noise in the rotation search will sometimes cause the cross-rotation peaks to deviate from their true NCS-consistent position. Empirically, we found that these values for the axis and angle thresholds allowed multiple NCS-consistent rotation sets to be identified in our test cases; if no NCS-consistent rotation sets are identified but confidence in the presence of NCS is high, the $\tau_{\rm axis}$ and $\tau_{\rm angle}$ thresholds may be relaxed.

For comparison, the translation-only RMSD (TRMSD) was computed between the model oriented according to each cross-rotation peak and each chain in the crystal structure. We define the TRMSD of two proteins to be the minimum main-chain C^α RMSD achievable when the molecules are only allowed to translate relative to one another (*i.e.* rotations are not allowed). Using this measure, we are able to quantify the similarity between models oriented by each cross-rotation peak and each monomer of the crystal structure. We emphasize that the TRMSD is not part of the *CRANS* algorithm but is rather a tool used to analyze the *CRANS* output to verify the correctness of our test cases.

Diffraction data for ChDHFR-TS were collected at Brookhaven National Laboratory (beamline X12C) as previously described (O'Neil *et al.*, 2003). Experiments were performed using three molecular-replacement models. The first model, LmDHFR-TS, consists of the DHFR-TS homodimer of the *Leishmania major* DHFR-TS protein (Knighton *et al.*, 1994). The second model, PcTSA, consists of the TS homodimer of the *Pneumocystis carinii* TS protein (PDB code 1f28; Anderson *et al.*, 2001), while the third model, PcTSB, is simply a rotated version of PcTSA with the flexible loop Asn186–Glu191 removed. Both the PcTSA and PcTSB models only consist of the TS homodimer, which represents approximately 60% of the entire DHFR-TS homodimer.

3. Results

The *CRANS* algorithm was tested on structure-factor data from four different protein crystals exhibiting three-, five- and sevenfold NCS. Three of these systems exhibit planar NCS, whereas ChDHFR-TS exhibits a 5_1 screw NCS (Fig. 3).

3.1. 2-Keto-3-deoxy-6-phosphogluconate aldolase (threefold NCS)

Structure factors for 2-keto-3-deoxy-6-phosphogluconate aldolase (1fq0) were obtained from the PDB. The first chain of 1fq0 was used as a model in a cross-rotation search using default parameters (§2.3). The resulting cross-rotation peak list was sorted by rotation-function score and contained 162 rotations. The TRMSD (§2.3) was computed between the model oriented by each cross-rotation peak and each chain of the 1fq0 crystal structure. While the average and standard deviation (in parentheses) TRMSD between each cross-rotation peak oriented model and the closest chain of the refined structure is 16.4 (7.2) Å, the closest (*i.e.* smallest TRMSD) peak to chain *A* has a TRMSD of 0.01 Å (peak 1), to chain *B* has a TRMSD of 4.00 Å (peak 137) and to chain *C* has a TRMSD of 3.44 Å (peak 46). In this test case, although peak 1 corresponds to the orientation closest to chain *A*, the top three rotation-function-ranked cross-rotation peaks do not correspond to correct model orientations. The *CRANS* algorithm was directed to search for complete threefold NCS (no missing peaks) and identified peaks 1 (chain *A*, TRMSD 0.01 Å), 159 (chain *B*, TRMSD 6.28 Å) and 59 (chain *C*, TRMSD 4.00 Å) as the highest scoring NCS-consistent rotation set (Table 1). While the NCS-consistent rotation set consisting of peaks 1, 137 and 46 contains those cross-rotation peaks producing orientations with the smallest TRMSD to the refined structure, these orientations do not strictly obey the NCS as tightly as peaks 1, 159 and 59 (originally found by *CRANS*). The orientations specified by cross-rotation peaks 1 and 137 have a relative angle of 108.6° rather than the NCS-specified 120°. Therefore, as a control, we relaxed the values of $\tau_{\rm axis}$ and $\tau_{\rm angle}$ and reran *CRANS*. With the relaxed thresholds, *CRANS* was able to find the NCS-consistent rotation set consisting of peaks 1, 137 and 46. Despite the fact that the overall quality of the rotations returned by the 1fq0 cross-rotation function are low compared with the cross-rotation peaks found in the other test cases (1chp, 1g31 and ChDHFR-TS), *CRANS* was still able to extract three rotations that result in properly oriented models.

The column 'Best Pk. in top 10' of Table 1 lists the cross-rotation peak with the smallest (best) TRMSD among the top ten rotation-function-ranked cross-rotation peaks. This column lists the peak which might be found in a manual molecular-replacement effort. Because *CRANS* exhaustively

checks all peaks in the cross-rotation list, NCS-consistent rotation peaks that appear at the bottom of the cross-rotation list are found as easily as those that appear near the top of the list.

3.2. Cholera toxin B subunit mutant (fivefold NCS)

Structure factors for cholera toxin B subunit mutant (1chp) were obtained from the PDB. The first chain of 1chp was used as a model in a cross-rotation search using default parameters. The resulting cross-rotation peak list was sorted by rotation-function score and contained 122 rotations. The TRMSD was computed between the model oriented by each cross-rotation peak and each chain of the 1chp crystal structure. In this case, unlike 1fq0, the top five cross-rotation peaks did have the lowest TRMSD to each of the five chains in the crystal structure, 0.10, 0.62, 1.02, 0.68 and 0.54 Å. The *CRANS* algorithm was directed to search for complete fivefold NCS (no missing peaks) and identified two sets of cross-rotation peaks. The first set contained peaks 1, 2, 3, 4 and 5, which correspond to the peaks with the lowest TRMSD to the refined structure (Table 1). *CRANS* also identified a second complete set of cross-rotation peaks containing peaks 59 (chain *D*, TRMSD 1.96 Å), 75 (chain *E*, TRMSD 2.83 Å), 58 (chain *F*, TRMSD 2.83 Å), 22 (chain *G*, TRMSD 1.91 Å) and 45 (chain *H*, TRMSD 1.97 Å). Thus, both identified rotation sets are consistent with the crystal structure.

3.3. Gp31 co-chaperonin from bacteriophage T4 (sevenfold NCS)

Structure factors for Gp31 co-chaperonin from bacteriophage T4 (1g31) were obtained from the PDB. The first chain of 1g31 was used as a model in a cross-rotation search using default parameters. The resulting cross-rotation peak list was sorted by rotation-function score and contained 47 rotations. The TRMSD was computed between the model oriented by each cross-rotation peak and each chain of the 1g31 crystal structure. The peaks with the lowest TRMSD to each crystallographic chain are listed in Table 1. High-degree NCS pushes the limits of standard molecular-replacement methods (Oh, 1995) since in these cases the model corresponds to a smaller percentage of the molecular mass in the unit cell (*e.g.* only 14.3% for sevenfold NCS). Therefore, it becomes increasingly likely that one or more NCS-consistent rotations will be missing from the cross-rotation peak list. In these cases, the ability of *CRANS* to generate missing peaks becomes rather useful. Thus, it is not surprising that a simple cross-rotation search for 1g31 performed with default search parameters is unable to identify rotations specifying orientations similar to chains *D* and *E* (*i.e.* orientations with low TRMSDs; Table 1). In this case, we would not expect the *CRANS* algorithm to find NCS-consistent rotation sets with less than two missing peaks. Note that the most complete NCS-consistent set found by the *CRANS* algorithm has three missing peaks (Table 1), indicating that although peak 8 has a TRMSD of 2.31 Å to chain *G*, the rotation differences between peak 8 and the other selected peaks did not satisfy the axis and angle thresholds τ_{axis} and τ_{angle}. The complete NCS-consistent rotation set generated by *CRANS* is consistent with the 1g31 crystal structure. The three missing peaks (generated by *CRANS*) have TRMSDs of 2.29, 2.29 and 2.37 Å (Table 1), thus demonstrating the ability of the *CRANS* algorithm to compute correct and complete NCS-consistent rotation sets even in the presence of missing cross-rotation peaks.

Table 1
TRMSDs measured for the three test systems 1fq0, 1chp and 1g31.

Column 2 lists the PDB chain identifier used in computing the TRMSD for the specified row. The 'Best Pk.' column lists the minimum TRMSD observed between the specified chain and all cross-rotation peaks, along with its corresponding cross-rotation peak index (in parentheses). Cross-rotation peaks are ordered by sorting them based on the cross-rotation function score, where a lower index corresponds to a higher (better) cross-rotation score. The TRMSD and peak index of the peak with the smallest TRMSD among the top ten cross-rotation function ranked peaks is listed in the 'Best Pk. in top 10' column. The TRMSD of the *CRANS*-identified peak of the NCS-consistent rotation set corresponding to the specified chain and its cross-rotation peak index (in parentheses) is listed in column '*CRANS* Pk.'. Peaks computed by *CRANS* (*i.e.* those missing in the cross-rotation peak list) are shown in italics with a dash for the peak index.

System	Chain	Best Pk.	Best Pk. in top 10	*CRANS* Pk.
1fq0	*A*	0.01 (1)	0.01 (1)	0.01 (1)
	B	4.00 (137)	14.43 (6)	6.28 (159)
	C	3.44 (46)	16.56 (3)	4.00 (59)
1chp	*D*	0.10 (1)	0.10 (1)	0.10 (1)
	E	0.62 (4)	0.62 (4)	0.62 (4)
	F	1.02 (3)	1.02 (3)	1.02 (3)
	G	0.68 (5)	0.68 (5)	0.68 (5)
	H	0.54 (2)	0.54 (2)	0.54 (2)
1g31	*A*	2.33 (2)	2.33 (2)	2.49 (3)
	B	2.37 (6)	2.37 (6)	2.37 (6)
	C	2.49 (4)	2.49 (4)	2.49 (4)
	D	8.13 (42)	10.64 (9)	*2.29 (—)*
	E	7.04 (1)	7.04 (1)	*2.29 (—)*
	F	2.29 (1)	2.29 (1)	2.29 (1)
	G	2.31 (8)	2.31 (8)	*2.37 (—)*

3.4. *C. hominis* DHFR-TS (5_1-fold NCS)

Molecular replacement was used to determine initial phase angles for the structure of ChDHFR-TS. Diffraction data to 2.8 Å were collected at Brookhaven National Laboratory (beamline X12C) and processed into structure factors as previously described (O'Neil *et al.*, 2003). Analysis of self-rotation peaks indicated the presence of fivefold NCS. A cross-rotation search using default parameters was performed with each of three models (LmDHFR-TS, PcTSA and PcTSB; see §2.3). The resulting cross-rotation peak lists were sorted by rotation-function score and run through the *CRANS* algorithm (because each model was a homodimer, *CRANS* was run in homodimer mode). The LmDHFR-TS, PcTSA and PcTSB cross-rotation peak lists contained 45, 38 and 38 rotations, respectively. Cross-rotation peak analysis with *CRANS* was able to find fivefold NCS sets with one missing peak for LmDHFR-TS, no missing peaks for PcTSA and two missing peaks for PcTSB. The NCS axes computed for all *CRANS*-identified NCS-consistent rotation sets agreed with the axis identified by the self-rotation search. Models oriented according to the rotations of the complete PcTSA NCS-

Table 2
TRMSDs measured for the three homologous models used in solving the ChDHFR-TS structure.

Column 2 lists the identifier of the homodimer used in computing the TRMSD for the specified row. The 'Closest Pk.' and '*CRANS* Pk.' columns are as listed in the caption of Table 1. The 'Closest dimer-flipped Pk.' column is similar to the 'Closest Pk.' column; however, TRMSDs are computed between the dimer flip of the specified homodimer and each peak in the cross-rotation peak list. Because *CRANS* computes model rotations invariant to homodimer flips, the *CRANS*-computed TRMSDs are taken as the smaller of the TRMSD to the crystallographic dimer or its dimer flip.

Model	Dimer	Closest Pk.	Closest dimer-flipped Pk.	*CRANS* Pk.
LmDHFR-TS	*A*	24.92 (39)	0.90 (4)	0.90 (4)
	B	0.87 (1)	21.55 (14)	0.87 (1)
	C	0.85 (2)	0.93 (5)	0.85 (2)
	D	1.07 (3)	20.23 (32)	1.07 (3)
	E	23.79 (4)	21.61 (31)	*1.33 (—)*
PcTSA	*A*	30.52 (24)	1.14 (2)	1.14 (2)
	B	1.17 (1)	15.92 (22)	1.17 (1)
	C	13.60 (17)	1.14 (3)	1.14 (3)
	D	1.23 (5)	20.83 (13)	1.23 (5)
	E	1.11 (4)	15.76 (36)	1.11 (4)
PcTSB	*A*	21.34 (11)	18.47 (10)	*1.34 (—)*
	B	1.21 (1)	1.15 (2)	1.21 (1)
	C	1.19 (4)	1.15 (3)	1.15 (3)
	D	1.19 (5)	14.39 (8)	1.19 (5)
	E	26.63 (37)	21.24 (35)	*1.32 (—)*

consistent rotation set were positioned using a translation search (O'Neil *et al.*, 2003; Crowther & Blow, 1967). The initial *R* factor of 52% was refined to 22.5% ($R_{\rm free}$ = 24.5%). Refined ChDHFR-TS molecules have a non-crystallographic 5_1 axis (O'Neil *et al.*, 2003) (see Fig. 3).

Analysis performed after the structure determination clearly explains the *CRANS* results. The TRMSD was computed between the thymidylate-synthase (TS) domain of each model oriented by each cross-rotation peak and each TS homodimer (dimers *A*, *B*, *C*, *D* and *E*) of the refined ChDHFR-TS crystal structure. Although the DHFR domains were used in the LmDHFR-TS cross-rotation search, the DHFR domains were not used in computing the TRMSDs because of the significant difference in the refined DHFR orientations relative to the highly conserved TS homodimer. The TRMSDs are presented in Table 2.

3.4.1. LmDHFR-TS. TRMSD analysis of the cross-rotation search peaks show that while some peaks closely approximate the structures of homodimers *A*, *B*, *C* and *D*, no cross-rotation peaks returned were similar to homodimer *E* (Table 2). Despite the significant structural and relative orientational differences between the DHFR domains of LmDHFR-TS and ChDHFR-TS, four of five correct rotations were still found in the LmDHFR-TS cross-rotation search using default search parameters. Table 2 shows that peaks 2 and 5 provide redundant information and both correspond to homodimer *C* (and its dimer flip). Therefore, the naïve selection of the top 5 cross-rotation peaks in the hope that they corresponded to the five correct NCS-related rotations would result in only four of five correct orientations and one redundant orientation. While peaks 2 and 5 corresponded to homodimer *C* and its dimer flip, the only cross-rotation peaks corresponding to homodimers *A*, *B* and *D* matched either the dimer or its dimer flip (but not both). Consequently, the ability of the *CRANS* algorithm to handle homodimer models was vital in this analysis. Specifically, a search for rotation-difference subgroups that ignored the fact that the model is a homodimer would not have been able to identify all four NCS-related peaks contained in the cross-rotation peak list. The peaks identified by *CRANS*

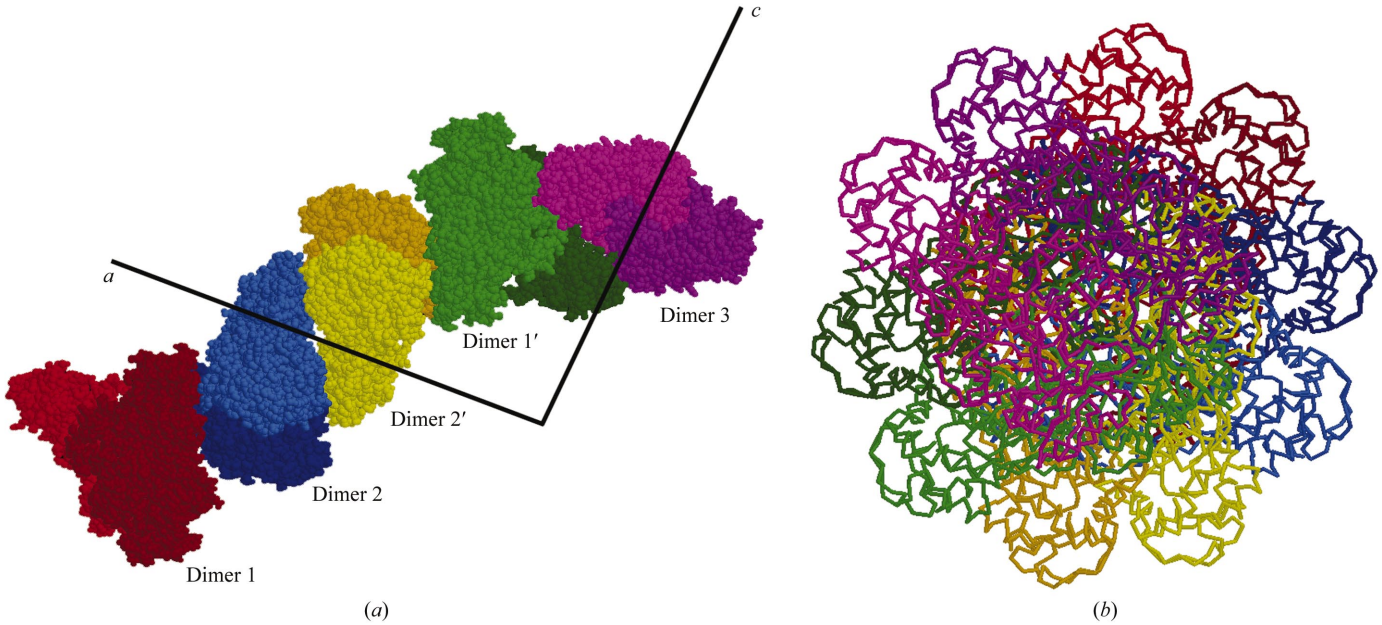

Figure 3
(*a*) A portion of the unit cell containing the 5_1 screw axis of ChDHFR-TS. The dimerization axis of dimer 3 is coincident with a crystallographic twofold. Dimers *x* and *x*′ are related by a crystallographic twofold. The crystallographic *a* and *c* unit-cell edges are shown for reference. (*b*) Top-down view showing the fivefold symmetry.

correspond to those peaks with the smallest TRMSD for each homodimer (Table 2). Furthermore, the rotation generated by *CRANS* to complete the NCS-consistent rotation set has a TRMSD to homodimer *E* of 1.33 Å.

3.4.2. PcTSA. Unlike the LmDHFR-TS model, PcTSA consists of only a TS homodimer. Cross-rotation peak analysis with *CRANS* found one fivefold NCS-consistent rotation set with no missing peaks (Table 2). The peaks identified by *CRANS* correspond to those with the smallest TRMSD to each of the ChDHFR-TS homodimers. As with LmDHFR-TS, the ability to handle homodimer models was crucial in *CRANS* analysis.

3.4.3. PcTSB. Similar to the PcTSA model, PcTSB consists of only a TS homodimer (see §2.3). Analysis of the cross-rotation peak sets with *CRANS* could not identify a fivefold NCS-consistent rotation sets with zero or one missing peaks; however, four NCS-consistent rotation sets were found with peaks corresponding to three of the five TS homodimers. Because peaks 1 and 2 and peaks 3 and 4 provide redundant information and both correspond to orientations of homodimer *B* and homodimer *C*, respectively, of the refined structure, the four identified sets correspond to the four sets {{1, 3, 5}, {2, 3, 5}, {1, 4, 5}, {2, 4, 5}} = {1, 2} × {3, 4} × {5} (numbers are cross-rotation peak numbers sorted by rotation-function score). The top scoring set was {1, 3, 5}, which has TRMSDs of 1.21, 1.15, 1.19 Å to the crystallographic ChDHFR-TS homodimers. The two computed (missing peaks) for this set have TRMSDs of 1.32 and 1.34 Å (Table 2). Although the top five cross-rotation peaks had a rotation-function score approximately twice that of the remaining peaks, the direct application of the top five peaks would not have resulted in the five orientations seen in the refined structure. Therefore, *CRANS* analysis provided the important information that not all five of the orientations seen in the refined structure were seen among the top five peaks of the cross-rotation list.

In summary, although two of the three cross-rotation peak lists did not contain all five NCS-consistent model rotations, the *CRANS* algorithm was successfully able to (i) verify that fivefold NCS was present, (ii) find sets of cross-rotation peaks related by an NCS-axis consistent with the self-rotation function for all three models and (iii) compute missing cross-rotation peaks corresponding to orientations with TRMSDs of 1.33, 1.34 and 1.32 Å to the final crystal structure. The identified NCS-consistent rotation sets were then used in a translation-function and rigid-body refinement to generate initial phase angles and the final structure of ChDHFR-TS (PDB code 1qzf; O'Neil *et al.*, 2003). The *CRANS*-identified NCS-consistent rotation sets were correct and agree with the final refined structure.

4. Discussion

In all six cross-rotation function searches performed only the default search parameters were used. That is, we did not spend any time optimizing cross-rotation search parameters. It is possible that by tweaking cross-rotation function search parameters and by optimizing the model (*i.e.* removing flexible loops, changing residues to Ala *etc*.) more NCS-consistent rotations could have been returned by the cross-rotation function. The time required to perform these optimizations can be reduced or eliminated by using *CRANS* to analyze the output of a cross-rotation search run with default parameters.

As the number of peaks returned by the cross-rotation function increases, the probability that these peaks will conspire to form 'fake' low-degree NCS-consistent rotation sets rises. This is especially true if the axis and angle tolerances are not particularly tight. Therefore, when analyzing low-degree NCS with large cross-rotation peak lists, the results of a *CRANS* search should be treated as a working hypothesis. Confidence in *CRANS*-identified NCS-consistent rotation sets can be increased by directing *CRANS* to use the NCS axis identified by the self-rotation map.

CRANS identifies NCS-consistent rotation sets from a cross-rotation search using monomer or homodimer models, since these are the most common model types used in molecular replacement; however, the *CRANS* algorithm can be extended to handle any degree of oligomerization (*e.g.* models that are homotrimers, homotetramers or homopentamers). Conceptually, to handle higher order model symmetry all symmetry rotations of the model are computed and then applied to each cross-rotation peak. For example, with *d*-fold model symmetry (*d*-fold oligomerization), let *f* be the $360/d^{\circ}$ rotation around the symmetry axis. Each rotation *r* identified by the cross-rotation function is replaced by the *d* rotations rf^{z} ($z \in \{0, 1, \ldots, d-1\}$). Heterodimers do not inherently contain a symmetry axis and therefore avoid the rotational degeneracy presented by homodimers. Thus, any model which does not have a symmetry axis (*e.g.* heterodimers, heterotrimers) should be treated as a monomer by *CRANS*.

While the tests in this paper were performed with an ordinary cross-rotation function, an alternative is to use a locked cross-rotation function (Tong & Rossmann, 1990; Tong, 2001) when the NCS axis is clear from the self-rotation function. Although the scoring of the rotation-function peaks is different, in the presence of NCS both the ordinary and locked cross-rotation functions should identify rotations corresponding to each of the *n* NCS-consistent orientations. When using a locked rotation function, non-crystallographic symmetry mates are generated using the NCS axis initially identified by the self-rotation function. However, if the NCS axis cannot be defined with a high degree of accuracy or the conservation of non-crystallographic symmetry in the crystal is not perfect, then a more accurate orientation for each monomer may be identified by orienting each monomer individually. *CRANS* can process the results of ordinary or locked cross-rotation functions to return a set of NCS-consistent orientations. The NCS-consistent model orientations identified and generated by *CRANS* can be positioned either sequentially using an ordinary translation function (Crowther & Blow, 1967) or simultaneously using a locked translation function (Tong & Rossmann, 1990; Tong, 2001) followed by an ordinary translation function (Tong, 2001).

5. Limitations and extensions

In this section, we discuss use of the *CRANS* algorithm in three situations: (i) crystals with improper NCS, (ii) crystals with multiple intersecting proper NCS axes and (iii) crystals with parallel NCS and crystallographic axes.

(i) The orientations of molecules involved in n-fold proper NCS are related by rotations of magnitude $360t/n$ ($t \in \{0, 1, \ldots, n-1\}$) around the NCS axis. By clustering the set of rotation differences between all pairs of cross-rotation peaks, *CRANS* is able to identify proper NCS-consistent rotation sets. Because there are no *a priori* constraints on the relative orientations between molecules with improper NCS, *CRANS* is unable to identify peaks whose orientations are only related by improper NCS. However, when a crystal possesses a combination of proper and improper NCS, the *CRANS* algorithm should identify the proper NCS, leaving those copies of the model only related by improper NCS to be positioned using alternative means. For example, in this situation, the Matthews coefficient (Matthews, 1968) might indicate the presence of k copies of the molecule per unit cell, while the self-rotation function indicates the presence of only n-fold proper NCS ($n < k$). If *CRANS* is able to identify one or more NCS-consistent rotation sets involving n molecules then one might assume that the remaining $k - n$ molecules are related by improper NCS. Alternatively, *CRANS* could be run in its scan mode to look for proper NCS with degree from 2 to k. In this scenario, *CRANS* might identify the presence of n-fold NCS, leaving the remaining $k - n$ molecules to be positioned by other techniques.

(ii) The identification of model orientations for crystals with multiple intersecting non-crystallographic rotation axes (*i.e.* 532 or 72 symmetry) is a challenge. The *CRANS* algorithm can be used to assist in the determination of correct orientations by identifying the cross-rotation peaks involved in each point group. *CRANS* will output all NCS-consistent rotation sets satisfying the $\tau_{\rm angle}$ and $\tau_{\rm axis}$ thresholds. Therefore, if there are multiple point-group symmetries satisfying these thresholds they will all be output by the *CRANS* algorithm. For example, in the case of 72 symmetry, *CRANS* could be directed to identify sevenfold proper NCS, which would identify two sevenfold NCS-consistent rotation sets with parallel axes. In the same situation, if *CRANS* were instead directed to search for twofold proper NCS, it would identify seven twofold proper NCS-consistent rotation sets with parallel axes. By independently identifying each NCS-consistent rotation set, *CRANS* could provide supporting evidence for the presence of multiple non-crystallographic rotation axes.

(iii) When an NCS axis is parallel to the crystallographic axis, cross-rotation peaks corresponding to NCS-consistent orientations may be buried under the crystallographic symmetry peaks. When this occurs, the cross-rotation function may have difficulty identifying accurate NCS-consistent orientations of the model. Because *CRANS* relies on the output of a cross-rotation function, it is limited by the quality of the identified cross-rotation peaks. Therefore, we would expect *CRANS* to have difficulty identifying NCS-consistent rotation sets with parallel crystallographic and non-crystallographic axes.

6. Conclusions

In this paper, we have presented the *CRANS* algorithm for analyzing lists of cross-rotation peaks both to extract NCS-consistent rotation sets and to complete partial sets by computing missing NCS-consistent rotations. We showed that the problem of identifying NCS-consistent rotation sets is equivalent to subgroup identification among differences in rotations of the cross-rotation peak list. We then tested the algorithm on four test proteins displaying three-, five- and sevenfold proper NCS using six models for molecular replacement (three from the solved crystal structure itself and three from homologous proteins). We demonstrated the ability of the *CRANS* algorithm to find NCS-consistent rotation sets both when all appropriate rotations were present and when up to three rotations (in the case of 1g31) were missing. Furthermore, *CRANS* successfully identified orientations that were used to generate initial phases in solving the structure of ChDHFR-TS. For all test cases, the *CRANS* algorithm was able to successfully generate correct NCS-consistent rotation sets. The *CRANS* algorithm is efficient, requiring only seconds on an Athlon-based processor to search for up to eightfold proper NCS on lists of 120 cross-rotation peaks.

By extracting more information from each cross-rotation peak list, the *CRANS* algorithm provides two main benefits to the crystallographer. Firstly, in the case where the cross-rotation peak list contains n peaks with significantly higher scores than the remainder of the list (where n is the NCS degree), the *CRANS* algorithm can confirm that these top peaks are indeed consistent with known proper NCS. The importance of this confirmation was demonstrated in the 1fq0, 1g31 and ChDHFR-TS systems, where the top n peaks did not correspond to the correct NCS-consistent orientations despite a sharp dropoff in rotation-function score after these top peaks. Secondly, when the model is partial or the NCS degree is high, it becomes likely that one or more NCS-consistent rotations will not be present in the cross-rotation peak list. When this happens, it is not possible to find a complete list of NCS-consistent cross-rotation peaks. The *CRANS* algorithm can find partial proper NCS-consistent rotation sets and then generate the missing proper NCS-consistent rotations to create a complete set. The ability to correctly generate missing peaks was demonstrated in the 1g31, LmDHFR-TS and PcTSB cases. By using quaternions to generate missing rotations, we avoid the rotational instability (*i.e.* singularities) that can arise when using other rotation representations.

7. Supporting materials

The *CRANS* program is distributed as a java jar file and is available at http://www.cs.dartmouth.edu/brd/Bio and by contacting the authors. The software is distributed under the Gnu Public License (Gnu, 2002).

APPENDIX *A*
Subgroup search

In this appendix, we show that the search for finite subgroups of $SO(3)$ (the special orthogonal group of three-dimensional rotations) among the rotation differences of cross-rotation peaks provably identifies NCS-consistent rotation sets. The *rotation difference set* of a set of n rotations R is defined as

$$D(R) = \{r_i^{-1} r_j \mid r_i, r_j \in R\}, \quad (2)$$

(where equal rotations are identified). Recall that $r_i^{-1} r_j$ is the rotation difference d_{ij} between rotations r_i and r_j. When $D(R)$ is a subgroup of $SO(3)$, we say that R and $D(R)$ are *complete*. When R is complete, a *partial rotation difference set* is a subset $D(R') \subset D(R)$ constructed from a subset of rotations R' of R. Therefore, a partial rotation-difference set can be made complete by adding missing rotation differences to complete the subgroup.

The goal of our algorithm is the following: given a set of cross-rotation peaks $C \subset SO(3)$, find finite subsets of rotations $R \subset C$ such that $D(R)$ is a subgroup of $SO(3)$.

We now show that an NCS-consistent rotation set R will have a complete rotation-difference set $D(R)$ satisfying the subgroup axioms and, conversely, that a rotation set R with a complete rotation-difference set $D(R)$ that satisfies the subgroup axioms is an NCS-consistent cross-rotation peak set.

For $D(R)$ to be a subgroup, the four group properties of identity, inverse, associativity and closure must be satisfied. All complete rotation difference sets satisfy the properties of identity, inverse and associativity. By definition (2 the rotation difference $d_{ii} = I$ (where I is the identity element) is present in all complete rotation-difference sets. The inverse property for each rotation difference d_{ij} from r_i to r_j is satisfied by the rotation from r_j to r_i; that is, $d_{ij} d_{ji} = I$. Finally, all rotations compose associatively by construction. Thus, the crux of the proof is in the closure property. To prove the forward direction, we note that all rotation differences are a multiple of $360/n^\circ$ around a common axis (where n is the degree of NCS) and that all n rotations $360t/n$ ($t \in \{0, 1, \ldots, n-1\}$) are represented in the set $D(R)$. Thus, the application of any number of rotation differences results in a rotation that (modulo 360°) is already in the complete rotation-difference set. Conversely, to prove the reverse direction, all finite subgroups have an identity and satisfy the closure property (modulo 360°); therefore, if they contain n members they represent n-fold symmetry and the symmetry angles $360t/n$ ($t \in \{0,1, \ldots, n-1\}$). This completes the proof.

When the model is a homodimer, each cross-rotation peak and its corresponding dimer-related symmetry mate produce equivalently oriented dimers. Therefore, each cross-rotation peak and its symmetry mate must be used to form finite subgroups and subsequently NCS-consistent rotation sets.

We now present a synthetic two-dimensional threefold NCS example to illustrate the framework presented in this section. Fig. 1(*a*) shows the model and five resulting cross-rotation function peaks $C = \{r_1, r_2, r_3, r_4, r_5\}$. The set of peaks $R = \{r_1, r_3, r_4\}$ forms an NCS-consistent rotation set. The complete rotation difference set $D(R)$ contains three rotations around an NCS-axis coming out of the page, d_α corresponds to the identity rotation, d_β corresponds to a 120° clockwise or 240° counterclockwise rotation and d_γ corresponds to a 240° clockwise or a 120° counterclockwise rotation. The subgroup $D(R)$ satisfies the four group properties (see Fig. 1*a*): the group property of associativity is satisfied by all rotations, d_α is the identity rotation and also serves as its own inverse, while d_β is the inverse of d_γ and d_γ is the inverse of d_β. Finally, the closure property is satisfied, $d_\alpha d_\beta = d_\beta d_\alpha = d_\beta$, $d_\alpha d_\gamma = d_\gamma d_\alpha = d_\gamma$ and $d_\beta d_\gamma = d_\gamma d_\beta = d_\alpha$.

APPENDIX *B*
CRANS input

The CRANS input file is a list of cross-rotation peaks in the *zxz* Euler angle rotation convention. In the *zxz* Euler angle format, the first angle specifies the rotation around the z axis, the second angle specifies the rotation around the new x axis and the third angle specifies the rotation around the new z axis. *CRANS* requires that the cross-rotation peak list have five columns corresponding to the peak index, z rotation, x rotation, z rotation and rotation-function score. Because this is the format used by *CNS* (Brünger *et al.*, 1998), the peak-list file generated by a *CNS* cross-rotation search can be used directly by the *CRANS* program.

This work is supported by grants to BRD from the National Institutes of Health (GM-65982) and the National Science Foundation (IIS-9906790, EIA-0102710, EIA-0102712, EIA-9818299, EIA-9802068 and EIA-0305444) and to ACA from the NIH (GM-067542). We thank Ramgopal Mettu, Robert O'Neil, Anthony Yan, Robert Stroud and all members of the Donald and Anderson Labs for helpful discussions and comments on drafts.

References

Anderson, A., O'Neil, R., Surti, T. & Stroud, R. (2001). *Chem. Biol.* **8**, 445–457.
Berman, H., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T., Weissig, H., Shindyalov, I. & Bourne, P. (2000). *Nucleic Acids Res.* **28**, 235–242.
Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W, Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst.* D**54**, 905–921.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Crowther, R. A. & Blow, D. M. (1967). *Acta Cryst.* **23**, 544–548.
Evans, P. R. (2001). *Acta Cryst.* D**57**, 1355–1359.
Gnu (2002) *The GNU General Public License.* http://www.gnu.org/licenses/licenses.html.
Hamilton, W. (1969). *Elements of Quaternions*, 3rd ed. New York: Chelsea Publishing Co.
Hunt, J., van der Vies, S., Henry, L. & Deisenhofer, J. (1997). *Cell*, **90**, 361–371.
Knighton, D., Kan, C., Howland, E., Janson, C., Hostomska, Z., Welsh, K. & Matthews, D. (1994). *Nature Struct. Biol.* **1**, 186–194.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
Merritt, E., Sarfaty, S., Chang, T., Palmer, L., Jobling, M., Holmes, R. & Hol, W. (1995). *Structure*, **3**, 561–570.

Oh, B.-H. (1995). *Acta Cryst.* D**51**, 140–144.
O'Neil, R., Lilien, R., Donald, B., Stroud, R. & Anderson, A. (2003). *J. Biol. Chem.* **278**, 52980–52987.
Rossmann, M. G. (1990). *Acta Cryst.* A**46**, 73–82.
Rossmann, M. G. & Blow, D. M. (1962). *Acta Cryst.* **15**, 24–31.
Salamin, E. (1979). *Applications of Quaternions to Computation with Rotations.* Technical Report, Stanford University, AI Lab. Internal working paper.
Tong, L. (2001). *Acta Cryst.* D**57**, 1383–1389.
Tong, L. & Rossmann, M. G. (1990). *Acta Cryst.* A**46**, 783–792.
Vagin, A. & Teplyakov, A. (1997). *J. Appl. Cryst.* **30**, 1022–1025.
Wymer, N., Buchanan, L., Henderson, D., Mehta, N., Botting, C., Pocivavsek, L., Fierke, C., Toone, E. & Naismith, J. (2001). *Structure*, **9**, 1–9.