Neutron Phasing Methods Applicable to D/H SIR Derivatives

David A. Langs, Hongliang Xu and Herbert A. Hauptman

Hauptman-Woodward Medical Research Institute and Department of Structural Biology,
SUNY at Buffalo, 700 Ellicott Street, Buffalo, NY 14203

Abstract: Twenty years ago Kossiakoff reviewed the then recent advances in neutron diffraction technology as applied to macromolecules and predicted that "Clearly the time is close at hand when neutron diffraction will take its place alongside X-ray diffraction as an equal partner in the structure determination of macromolecules." The first, and perhaps only successful application to determine a large crystal structure by SIR phasing methods occurred when Roger Koeppe enlisted the help of Beno Schoenborn to phase an orthorhombic form of gramicidin A via H/D exchange. A 5Å map based on 8 determined H/D replacement sites revealed the cylindrical shape of the dimmer shown in figure 1, but the helical trace of the peptide chain was not discernable. Koeppe also synthesized a deuterated Val[1] derivative to improve the phases by MIR methods, but crystals of sufficient size could not be obtained. In retrospect, it may have been impossible to determine the Val[1] replacement sites since H/D positions on the same methyl group would be only 1.7Å apart, and not resolved by the experimental resolution. In this regard we have developed an SIR structure solution method that does not require us to determine the replacement sites. Instead, low resolution macromolecular phases can be determined directly from *a priori* phase invariant triples estimates based on the SIR amplitudes.

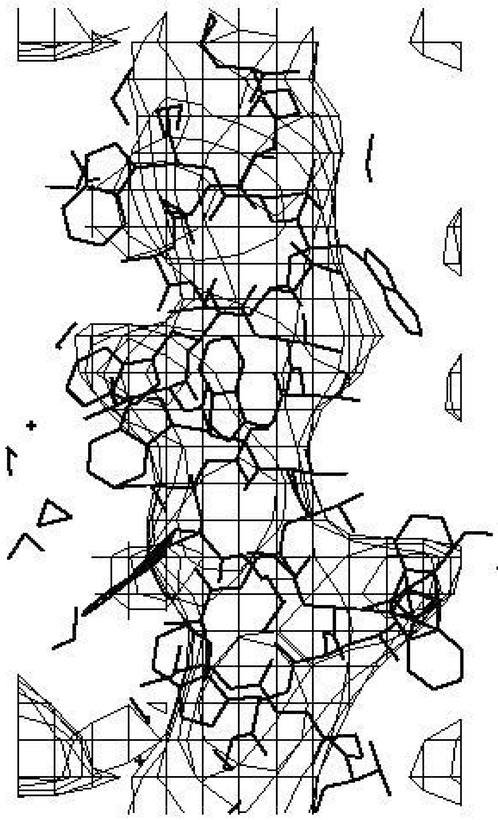

Figure 1

Negative Density & Direct Methods: For many years most crystallographers believed that map "positivity" was a necessary condition for direct methods validity even though Hauptman (1976) showed that direct methods could be rigorously extended to the case of unequal atoms, even to the extreme that some could have negative scattering density, such as H in neutron diffraction.. With this in mind our direct methods program **SnB** was modified to include the most negative peak positions as hydrogen atoms in structure factor calculations to recompute phases. Tests performed on experimental neutron data for cyclosporin A [$C_{62}H_{111}N_{11}O_{12}\cdot H_2O$, 86% complete to 0.915Å; Knott, Schefer & Schoenborn, 1990] gave a 2% NSnB success rate for this 199 atom structure in spite of the fact that only two of the 9000 triples used had A-values greater than 1.5.

Is Negative Density Harmful? : Error-free neutron data were used to compare the fully deuterated structure with the native hydrogenated one. Similarly, error-free X-ray data were tested for the same purpose. ***These results were contrary to our expectations***.

Cyclosporin	success rate	<|δφ|>	#atoms
NSnB			
(H-structure)	6 %	30°	199
(D-structure)	0.3 %	50°	199
X-RAY			
(SnB)	1.1 %	13°	86
(RANTAN)	0.1 %	25°	86

In view of this evidence, one is forced to conclude that ***negative density is an asset rather than a hindrance*** when it comes to direct methods. Tests indicate that structures as large as lysozyme can be solved by NSnB if data < 0.90Å are available. Larger structures, however, will require derivative SIR data to succeed at much lower resolution.

SIR Direct Methods: One tremendous advantage afforded by neutron diffraction is that an H/D substitution strategy can provide an enormous number of ***perfect*** isomorphous replacement derivatives. The same can not be said of X-ray data. Since estimates for the three-phase triples invariants for SIR data were formulated 20 years ago, it is a fairly straightforward proposition to incorporate these results into NSnB. Error-free neutron data for a pair of cyclosporin structures, S and T, was tested to this end.

$$S : C_{62}D_{111}N_{11}O_{12}\cdot D_2O; \quad T : C_{62}D_{106}H_5N_{11}O_{12}\cdot H_2O$$

SIR triples A-values were computed from the six E-magnitudes $S_{\mathbf{h}}$, $S_{\mathbf{k}}$, $S_{\mathbf{l}}$, $T_{\mathbf{h}}$, $T_{\mathbf{k}}$, $T_{\mathbf{l}}$ according to Hauptman's 1982 formula. The A_{SIR} values ranged from +10 to -10 with 25,000 triples exceeding the 1.5 value indicated for the native neutron data. These large A_{SIR} values indicate an enormous phasing power, such that phasing may be achieved at much lower resolution. In this regard we began with 0.915Å data, and if successful, proceeded to lower the resolution to determine at which point difficulties ensued. In this way we were able to solve the structure at as low as 2.27Å resolution, ***which bodes well for potential macromolecular applications using neutron data from spallation sources***.

The 2.27Å Results: 1000 random NSnB trials were each iterated for 150 cycles and sorted in ascending order on R_{min} as is the usual practice with X-ray data. Density modification [ρ=0 if ρ<0.25σ(ρ); ρ=2 σ(ρ) if ρ>2 σ(ρ)] replaced "peak picking" for the lower resolution ranges. 13 Solutions are noted among the 75 sets with the lowest R_{min} values (figure 2), but are not clearly separated from the non-solution on this basis.

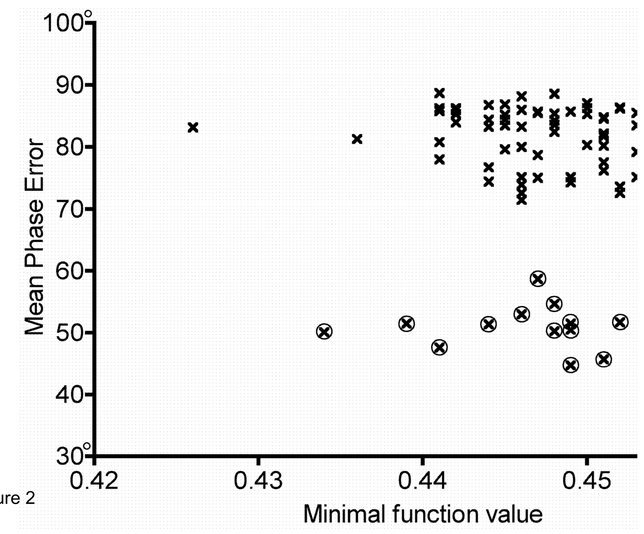

Figure 2

To better identify the 13 solutions a **sine invariant correlation coefficient** was proposed

$$CC(i,j) = [\ \langle \sin\Phi_{h,k} \sin\Psi_{h,k} \rangle - \langle \sin\Phi_{h,k} \rangle \langle \sin\Psi_{h,k} \rangle\ ] \,/\, \sigma(\sin\Phi_{h,k})\ \ \sigma(\ \sin\Psi_{h,k}\ )$$

where $\Phi_{h,k}$ and $\Psi_{h,k}$ are the triples sine invariants of the i^{th} and j^{th} sets, respectively. Large |CC(i,j)| values are expected for equivalent phase sets. The top 75 phase sets in Table 1 will generate 75*74/2 = 2775 CC(i,j) values. When these values are sorted in descending order on |CC| the topmost entries (Table 1) will form the basis for identifying the groups of phase sets that are equivalent except for the choice of origin and enantiomorph.

Table 1. Top 75 Pairwise comparisons ranked on CC

rank	amb1	amb2	CC	rank	amb1	amb2	CC	rank	amb1	amb2	CC
1	677*	1169*	0.224	26	1281*	1611*	0.177	51	553	567	0.160
2	33*	281*	0.221	27	1045*	1281*	0.177	52	117*	1281*	0.160
3	677*	1045*	0.218	28	553	1397	0.176	53	117*	1169*	0.160
4	655*	1169*	0.217	29	551	715	0.175	54	281*	1169*	0.158
5	31	961	0.212	30	553	1957	0.173	55	33*	1281*	0.158
6	117*	1015*	0.210	31	1059	1921	0.171	56	1711	1807	0.157
7	827	1305	0.209	32	65	595	0.170	57	1305	1921	0.157
8	281*	677*	0.209	33	281*	1281*	0.169	58	589	1059	0.156
9	371	1921	0.207	34	1397	1957	0.168	59	371	1059	0.156
10	677*	1281*	0.205	35	1365	1431	0.168	60	33*	677*	0.156
11	1169*	1611*	0.203	36	827	1921	0.168	61	677*	1611*	0.155
12	171	1717	0.203	37	33*	1875*	0.168	62	589	1345	0.155
13	117*	655*	0.203	38	117*	1611*	0.167	63	31	803	0.155
14	1397	1935	0.202	39	1677	1711	0.166	64	1305	1711	0.154
15	281*	1611*	0.192	40	589	1921	0.165	65	715	1365	0.152
16	1717	1935	0.191	41	1045*	1169*	0.163	66	595	827	0.152
17	715	1431	0.189	42	1015*	1281*	0.162	67	33*	1045*	0.152
18	371	1691	0.189	43	677*	1015*	0.162	68	1203	1983	0.151
19	1169*	1281*	0.188	44	371	1305	0.162	69	1169*	1737*	0.151
20	371	759	0.187	45	281*	655*	0.161	70	1397	1717	0.150
21	563	1711	0.186	46	209	1677	0.161	71	589	827	0.150
22	407	1711	0.182	47	181	1203	0.161	72	553	583	0.150
23	33*	1611*	0.182	48	1513	1853	0.160	73	551	827	0.150
24	371	827	0.179	49	1345	1711	0.160	74	65	1921	0.150
25	655*	1045*	0.178	50	583	1957	0.160	75	1939	1999	0.149

CLUSTER ANALYSIS: Only four equivalent groups of phases can be distilled from the |CC| data indicated in Table 1. The omit-map correlation coefficient and $\langle\rho^7\rangle$ of the electron density are slightly better than R_{min}. for identifying the correct solution.

Group	# of sets	<|δφ|>	R_{min}	$CC(X_{\mathbf{h}},F_{\mathbf{h}})$	$\langle\rho^7\rangle$
I	13	35°	0.423	0.581	322
II	5	71°	0.432	0.397	159
III	10	76°	0.437	0.510	167
IV	41	84°	0.407	0.271	72

When the 13 sets are phase shifted to the same origin and enantiomoph, and then phase averaged, the resulting set has a mean phase error of 38° and this 2.27Å map reveals the peptide backbone and side-chains of the structure shown in figure below.

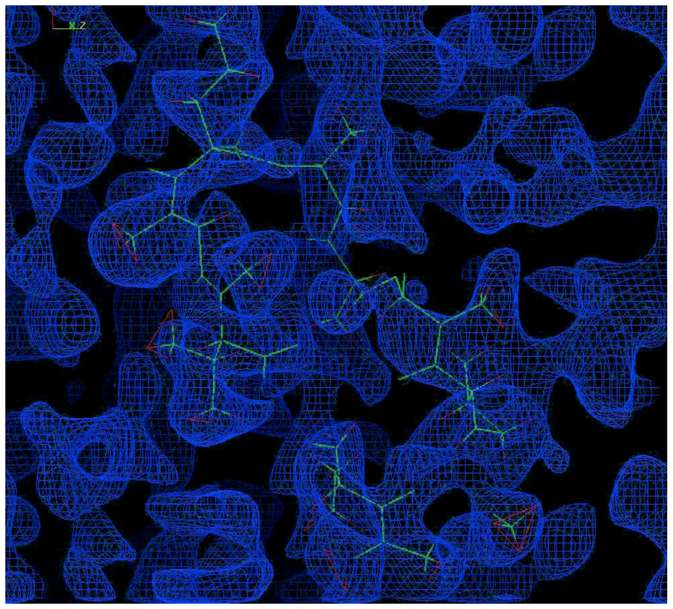

Figure 3

Larger Macromolecules: The neutron structure of a mutant rubredoxin [Chatake et al. *Acta Cryst* **D60**, 1364 (2004)] was tested using error-free SIR data at the published resolution of 1.6Å. The fully deuterated crystal was considered to be the native and SIR derivatives were considered by replacing certain deuterated amino acids in the 52 residue sequence. A Pro^5-mutant replaced the 5 proline residues in the sequence with wild type H-labeled proline. There were a total of 7426 data of which the 5500 largest E-values were phased by the NSnB method. 400 random trials were each run for 300 SnB cycles. CC values were computed for the 400*399/2 = 79800 pairs of phases sets. The 3rd largest CC value identified the phase sets that had the lowest mean phase errors, 68.8 and 76.3°. Additional derivatives would markedly improve

Amb1	<|δφ|>	Amb2	<|δφ|>	CC	
561	88.6	671	89.4	0.186	
611	88.9	661	88.4	0.172	
333	**68.8**	689	**76.3**	0.169	***
509	88.8	671	89.4	0.142	
509	88.8	561	88.6	0.141	
203	88.1	497	88.6	0.140	
37	88.3	493	88.3	0.136	
199	88.6	795	89.2	0.131	
51	88.6	489	88.9	0.131	
387	89.3	537	88.8	0.128	