

Automated de novo Electron Density Map Tracing for the Structural Genomics Era

Dipesh Risal

Accelrys Inc., 10188 Telesis Court, Suite 100, San Diego, CA 92121, USA

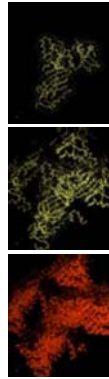


Abstract Structural genomics initiatives around the world have gained momentum in recent years. An early step in such projects is the tracing of initial electron density maps, and this remains a challenging step requiring significant expertise. Map interpretation is especially demanding at lower resolutions, or when there are errors associated with the phase information. Two methods of tracing electron density maps have been implemented in the crytographic modeling environment QUANTA. The first method is optimal for high-resolution (2.0 Å) data sets, and involves simultaneous multiple path analysis to identify the optimal path of the protein chain in such terms and loops. The second method improves on the limitations of existing auto-tracing programs by extending the effective resolution limits from 2.2 Å to >4.0 Å. Here we present results of the two tracing methods when applied to datasets with different resolution limits and figures of merit. The resultant alpha carbon traces, as well as all atom models (built with QUANTA), are compared to the respective published structures. The two methods are extremely robust and fast (a few seconds for the high-resolution tracing, and a few minutes for the low-resolution tracing), and can trace the majority of alpha carbons in electron density maps with figures of merit as low as 0.5.

Introduction The many active structural genomics initiatives worldwide have started to produce a large number of new protein structures. The majority of these new structures is solved by de novo methods such as X-ray crystallography, cryo-electron microscopy, and NMR spectroscopy. However, the high-resolution limit of X-ray crystallography is being pushed to lower resolutions, and the high-resolution limit of cryo-EM is being pushed to higher resolutions. This has led to a significant increase in the number of structures solved at lower resolutions. The second method improves on the limitations of existing auto-tracing programs by extending the effective resolution limits from 2.2 Å to >4.0 Å. Here we present results of the two tracing methods when applied to datasets with different resolution limits and figures of merit. The resultant alpha carbon traces, as well as all atom models (built with QUANTA), are compared to the respective published structures. The two methods are extremely robust and fast (a few seconds for the high-resolution tracing, and a few minutes for the low-resolution tracing), and can trace the majority of alpha carbons in electron density maps with figures of merit as low as 0.5.

Existing software programs such as ARP-wARP, MAD and RESOLVE carry out automated de novo tracing and map interpretation. However, these programs are designed to trace and interpret maps at high resolutions (differences between 2.0 and 4.0 Å). However, data sets at lower resolutions do not fare as well with these programs. Furthermore, the programs take several hours of CPU time to run to completion (1). For data in the range 3 Å to 2.1 Å, initial tracing of the C α chain is not the rate-limiting step for structure determination. However, this step requires expert knowledge and a significant amount of user time, since there is a large number of possible paths through the density map. The QUANTA software suite includes an automated algorithm that has been introduced in QUANTA 2005 (2). The first is geared towards high resolution data between 2.2 Å and 1.1 Å) and typically runs to completion in 1 or 2 seconds. The second algorithm can reliably trace most of the alpha carbons in electron density maps at lower resolutions (3.0 Å to 4.0 Å). The second algorithm is based on an all atom model and real-space refined with the automated sequence assignment and model rebuilding tools available in QUANTA. As such, these tools represent an extremely fast and complete package for de novo structure determination for both "test tube" data, and high resolution data sets.

Functionality The two tracing methods require an electron density map. Symmetry information is helpful in determining the extents of the asymmetric unit (ASU), but not necessary. The map must be skeletonized into bones atoms, with an appropriate sigma level. A mask from a density modification experiment can be used to mask the map. Alternatively, the bones atoms must be edited to represent the ASU, and a mask created from the edited bones.



High Resolution Tracing Electron density maps with resolution worse than 3.1 Å generally have connected electron density. The high resolution tracing is designed to work with electron density maps with connected electron density, as well as the presence of peaks in approximate C α positions. The tracing method looks for all the possible paths through the density map. The tracing method uses a correlated pathway analysis to work out the ideal path from all the possible paths found from the bones.



Low Resolution Tracing At lower resolutions, there is generally continuous density along the main chain. The tracing protocol attempts to identify the trace in a two-step step process.

1) Identification of secondary structure elements

This has been described in detail elsewhere (3). Briefly, shape analysis of a bone path is performed to determine whether the path corresponds to an alpha or beta strand. Multiple vectors are generated with the aim of reducing false positives. The vectors are then clustered to remove degeneracy. Finally, vectors are rejected based on the principle that secondary structure elements do not overlap.



Case 4: (Ribonuclease SA) (11)

Resolution (Å)	Trace Quality	Trace Length (residues)	Trace Coverage (%)
2.2	0.5	100	100
2.5	0.5	100	100
3.0	0.5	100	100
3.5	0.5	100	100
4.0	0.5	100	100

Case 5: (Oxide Membrane Phospholipase A) (12)

Resolution (Å)	Trace Quality	Trace Length (residues)	Trace Coverage (%)
2.2	0.5	100	100
2.5	0.5	100	100
3.0	0.5	100	100
3.5	0.5	100	100
4.0	0.5	100	100

Case 6: (trans glycosylase) (13)

Resolution (Å)	Trace Quality	Trace Length (residues)	Trace Coverage (%)
2.2	0.5	100	100
2.5	0.5	100	100
3.0	0.5	100	100
3.5	0.5	100	100
4.0	0.5	100	100

Case 7: (RF1 RNA pseudo-uridine synthase) (7)

Resolution (Å)	Trace Quality	Trace Length (residues)	Trace Coverage (%)
2.2	0.5	100	100
2.5	0.5	100	100
3.0	0.5	100	100
3.5	0.5	100	100
4.0	0.5	100	100

Case 8: (RF2 RNA pseudo-uridine synthase) (7)

Resolution (Å)	Trace Quality	Trace Length (residues)	Trace Coverage (%)
2.2	0.5	100	100
2.5	0.5	100	100
3.0	0.5	100	100
3.5	0.5	100	100
4.0	0.5	100	100

Case 9: (RF3 RNA pseudo-uridine synthase) (7)

Resolution (Å)	Trace Quality	Trace Length (residues)	Trace Coverage (%)
2.2	0.5	100	100
2.5	0.5	100	100
3.0	0.5	100	100
3.5	0.5	100	100
4.0	0.5	100	100

Case 10: (RF4 RNA pseudo-uridine synthase) (7)

Resolution (Å)	Trace Quality	Trace Length (residues)	Trace Coverage (%)
2.2	0.5	100	100
2.5	0.5	100	100
3.0	0.5	100	100
3.5	0.5	100	100
4.0	0.5	100	100

Case 11: (RF5 RNA pseudo-uridine synthase) (7)

Resolution (Å)	Trace Quality	Trace Length (residues)	Trace Coverage (%)
2.2	0.5	100	100
2.5	0.5	100	100
3.0	0.5	100	100
3.5	0.5	100	100
4.0	0.5	100	100

Case 12: (RF6 RNA pseudo-uridine synthase) (7)

Resolution (Å)	Trace Quality	Trace Length (residues)	Trace Coverage (%)
2.2	0.5	100	100
2.5	0.5	100	100
3.0	0.5	100	100
3.5	0.5	100	100
4.0	0.5	100	100

Case 13: (RF7 RNA pseudo-uridine synthase) (7)

Resolution (Å)	Trace Quality	Trace Length (residues)	Trace Coverage (%)
2.2	0.5	100	100
2.5	0.5	100	100
3.0	0.5	100	100
3.5	0.5	100	100
4.0	0.5	100	100

Case 14: (RF8 RNA pseudo-uridine synthase) (7)

Resolution (Å)	Trace Quality	Trace Length (residues)	Trace Coverage (%)
2.2	0.5	100	100
2.5	0.5	100	100
3.0	0.5	100	100
3.5	0.5	100	100
4.0	0.5	100	100

Case 15: (RF9 RNA pseudo-uridine synthase) (7)

Resolution (Å)	Trace Quality	Trace Length (residues)	Trace Coverage (%)
2.2	0.5	100	100
2.5	0.5	100	100
3.0	0.5	100	100
3.5	0.5	100	100
4.0	0.5	100	100

Case 16: (RF10 RNA pseudo-uridine synthase) (7)

Resolution (Å)	Trace Quality	Trace Length (residues)	Trace Coverage (%)
2.2	0.5	100	100
2.5	0.5	100	100
3.0	0.5	100	100
3.5	0.5	100	100
4.0	0.5	100	100

Case 17: (RF11 RNA pseudo-uridine synthase) (7)

Resolution (Å)	Trace Quality	Trace Length (residues)	Trace Coverage (%)
2.2	0.5	100	100
2.5	0.5	100	100
3.0	0.5	100	100
3.5	0.5	100	100
4.0	0.5	100	100

Case 18: (RF12 RNA pseudo-uridine synthase) (7)

Resolution (Å)	Trace Quality	Trace Length (residues)	Trace Coverage (%)
2.2	0.5	100	100
2.5	0.5	100	100
3.0	0.5	100	100
3.5	0.5	100	100
4.0	0.5	100	100

Case 19: (RF13 RNA pseudo-uridine synthase) (7)

Resolution (Å)	Trace Quality	Trace Length (residues)	Trace Coverage (%)
2.2	0.5	100	100
2.5	0.5	100	100
3.0	0.5	100	100
3.5	0.5	100	100
4.0	0.5	100	100

Case 20: (RF14 RNA pseudo-uridine synthase) (7)

Resolution (Å)	Trace Quality	Trace Length (residues)	Trace Coverage (%)
2.2	0.5	100	100
2.5	0.5	100	100
3.0	0.5	100	100
3.5	0.5	100	100
4.0	0.5	100	100

Case 21: (RF15 RNA pseudo-uridine synthase) (7)

Resolution (Å)	Trace Quality	Trace Length (residues)	Trace Coverage (%)
2.2	0.5	100	100
2.5	0.5	100	100
3.0	0.5	100	100
3.5	0.5	100	100
4.0	0.5	100	100

Case 22: (RF16 RNA pseudo-uridine synthase) (7)

Resolution (Å)	Trace Quality	Trace Length (residues)	Trace Coverage (%)
2.2	0.5	100	100
2.5	0.5	100	100
3.0	0.5	100	100
3.5	0.5	100	100
4.0	0.5	100	100

Case 23: (RF17 RNA pseudo-uridine synthase) (7)

Resolution (Å)	Trace Quality	Trace Length (residues)	Trace Coverage (%)
2.2	0.5	100	100
2.5	0.5	100	100
3.0	0.5	100	100
3.5	0.5	100	100
4.0	0.5	100	100

Case 24: (RF18 RNA pseudo-uridine synthase) (7)

Resolution (Å)	Trace Quality	Trace Length (residues)	Trace Coverage (%)
2.2	0.5	100	100
2.5	0.5	100	100
3.0	0.5	100	100
3.5	0.5	100	100
4.0	0.5	100	100

Case 25: (RF19 RNA pseudo-uridine synthase) (7)

Resolution (Å)	Trace Quality	Trace Length (residues)	Trace Coverage (%)
2.2	0.5	100	100
2.5	0.5	100	100
3.0	0.5	100	100
3.5	0.5	100	100
4.0	0.5	100	100

Case 26: (RF20 RNA pseudo-uridine synthase) (7)

Resolution (Å)	Trace Quality	Trace Length (residues)	Trace Coverage (%)
2.2	0.5	100	100
2.5	0.5	100	100
3.0	0.5	100	100
3.5	0.5	100	100
4.0	0.5	100	100

Case 27: (RF21 RNA pseudo-uridine synthase) (7)

Resolution (Å)	Trace Quality	Trace Length (residues)	Trace Coverage (%)
2.2	0.5	100	100
2.5	0.5	100	100
3.0	0.5	100	100
3.5	0.5	100	100
4.0	0.5	100	100

Case 28: (RF22 RNA pseudo-uridine synthase) (7)

Resolution (Å)	Trace Quality	Trace Length (residues)	Trace Coverage (%)
2.2	0.5	100	100
2.5	0.5	100	100
3.0	0.5	100	100
3.5	0.5	100	100
4.0	0.5	100	100

Case 29: (RF23 RNA pseudo-uridine synthase) (7)

Resolution (Å)	Trace Quality	Trace Length (residues)	Trace Coverage (%)
2.2	0.5	100	100
2.5	0.5	100	100
3.0	0.5	100	100
3.5	0.5	100	100
4.0	0.5	100	100

Case 30: (RF24 RNA pseudo-uridine synthase) (7)

Resolution (Å)	Trace Quality	Trace Length (residues)	Trace Coverage (%)
2.2	0.5	100	100
2.5	0.5	100	100
3.0	0.5	100	100
3.5	0.5	100	100
4.0	0.5	100	100

Case 31: (RF25 RNA pseudo-uridine synthase) (7)

Resolution (Å)	Trace Quality	Trace Length (residues)	Trace Coverage (%)
2.2	0.5	100	100
2.5	0.5	100	100
3.0	0.5	100	100
3.5	0.5	100	100
4.0	0.5	100	100

Case 32: (RF26 RNA pseudo-uridine synthase) (7)

Resolution (Å)	Trace Quality	Trace Length (residues)	Trace Coverage (%)
2.2	0.5	100	100
2.5	0.5	100	100
3.0	0.5	100	100
3.5	0.5	100	100
4.0	0.5	100	100

Case 33: (RF27 RNA pseudo-uridine synthase) (7)

Resolution (Å)	Trace Quality	Trace Length (residues)	Trace Coverage (%)
2.2	0.5	100	100
2.5	0.5	100	100
3.0	0.5	100	100
3.5	0.5	100	100
4.0	0.5	100	100

Case 34: (RF28 RNA pseudo-uridine synthase) (7)

Resolution (Å)	Trace Quality	Trace Length (residues)	Trace Coverage (%)
2.2	0.5	100	100
2.5	0.5	100	100
3.0	0.5	100	100
3.5	0.5	100	100
4.0	0.5	100	100