Computer-Assisted Structural Elucidation: a Further Look at Two Natural Products

Gene Mazzola and Clark Ridge

UMD-FDA Joint Institute

UNIVERSITY OF MARYLAND

Spectral Data Requirements

1. ^{1}H NMR data: peak integral information
2. HSQC, COSY, HMBC
3. TOCSY – when COSY spectra indicate complex spin systems
4. ^{13}C NMR data: identify quaternary carbons; especially important for those carbons with no HMBC connectivities
5. HRMS – for MF determination
6. IR, UV/vis – information on functional groups

Format

1. FIDs and Processing Parameters
2. TXT Files: time consuming and transcription errors

Conditions Applied to NMR Data Processing

1. HMBC Correlations
   (a.) Strong: $^{2}J_{CH} - {}^{3}J_{CH}$
   (b.) Weak: $^{2}J_{CH} - {}^{4}J_{CH}$
2. Direct heteroatom-to-heteroatom connectivities disallowed
3. Triple bonds within 3-, 4-, and 5-membered rings disallowed

Generation of Molecular-Connectivity Diagrams

1. Heteronuclear correlations (HMBC)
2. Homonuclear correlations (COSY, possibly TOCSY) between CH_n fragments (where n = 1-3)

Phase 1 of Structural Determination

1. Generation of potential structures (tens to hundreds)

2. “Rough” calculation of ^{13}C chemical shifts for all structures

3. “Fast-deviation” statistic, $d_F(^{13}C) = \delta_{experimental} - \delta_{predicted}$

4. Potential structures are initially ranked in order of increasing $d_F(^{13}C)$

Phase 2 of Structural Determination

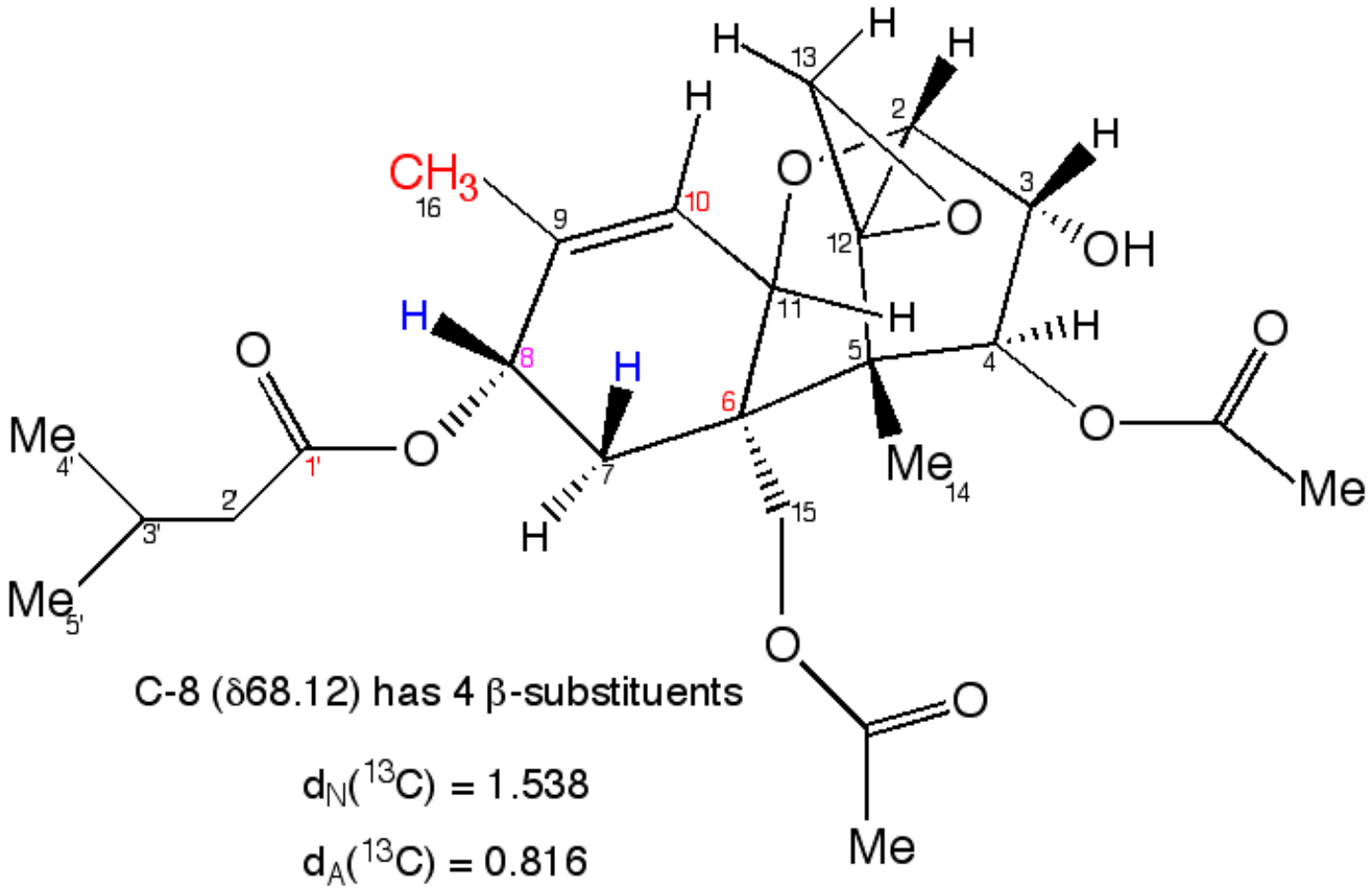
1. More accurate ^{13}C chemical shifts calculated for the smaller of:
   (a.) All structures with $d_F(^{13}C) \leq 4$ ppm/carbon (for a small no. of structures)
   (b.) The first 50 ranked structures (for a large number of structures)
2. Differences between more accurately predicted and experimental data are reported as one or both types of statistics:
   (a.) $d_A(^{13}C)$: HOSE (Hierarchically Ordered Spherical Description of Environment)
   (b.) $d_N(^{13}C)$: Neural Network
3. Potential structures are re-ranked in order of increasing $d_{A/N}(^{13}C)$ values
4. Structures having $d_{A/N}(^{13}C)$ values > 4 ppm/carbon are generally discarded
5. Correct structures are usually identified at this point
6. When the smallest $d_{A/N}(^{13}C)$ values are very similar, the following may be calculated: $d_{A/N}(^{1}H)$ and d(MS)

Two Natural Products Revisited

1. T-2 Toxin: 2 similar best structures
   (a.) $d_N(^{13}C)$ values: 1.538 and 2.395
   (b.) $d_A(^{13}C)$ values: 0.816 and 2.187

2. Kauradienoic Acid: 2 assignment schemes
   (a.) $d_N(^{13}C)$ values: 1.17 and 1.26

T-2 Toxin Correct Structure

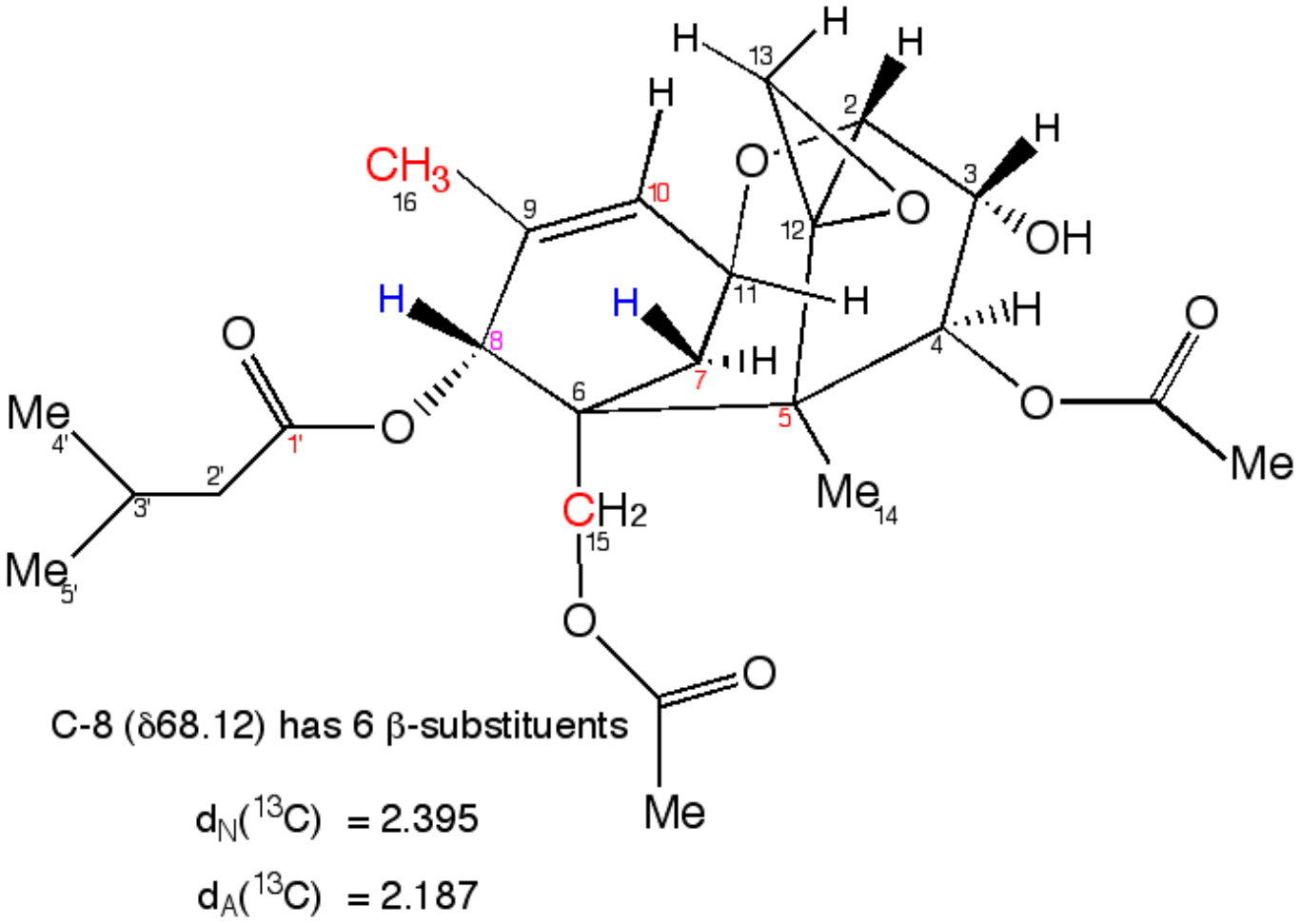


C-8 (δ68.12) has 4 β-substituents

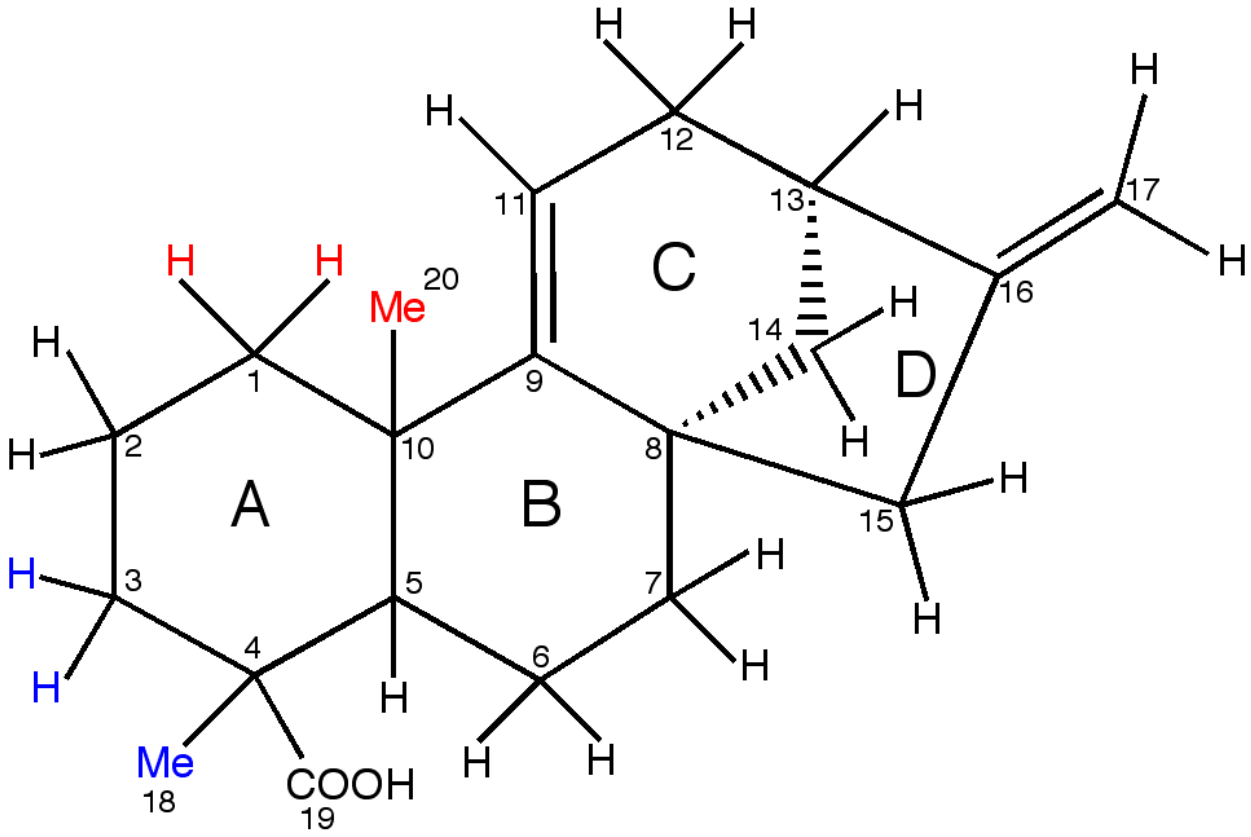
$d_N(^{13}C) = 1.538$

$d_A(^{13}C) = 0.816$

T-2 Toxin Incorrect Structure



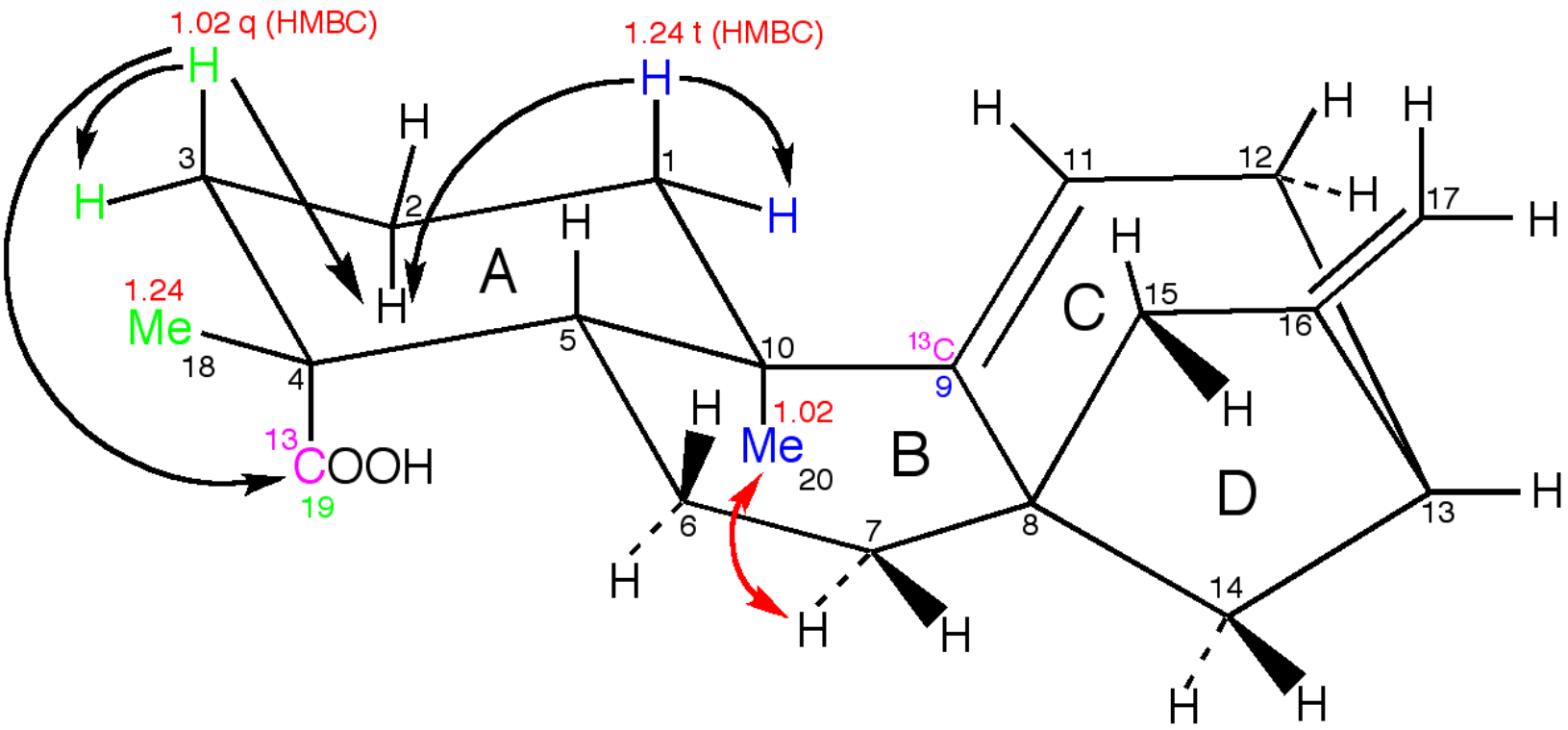
Kauradienoic Acid



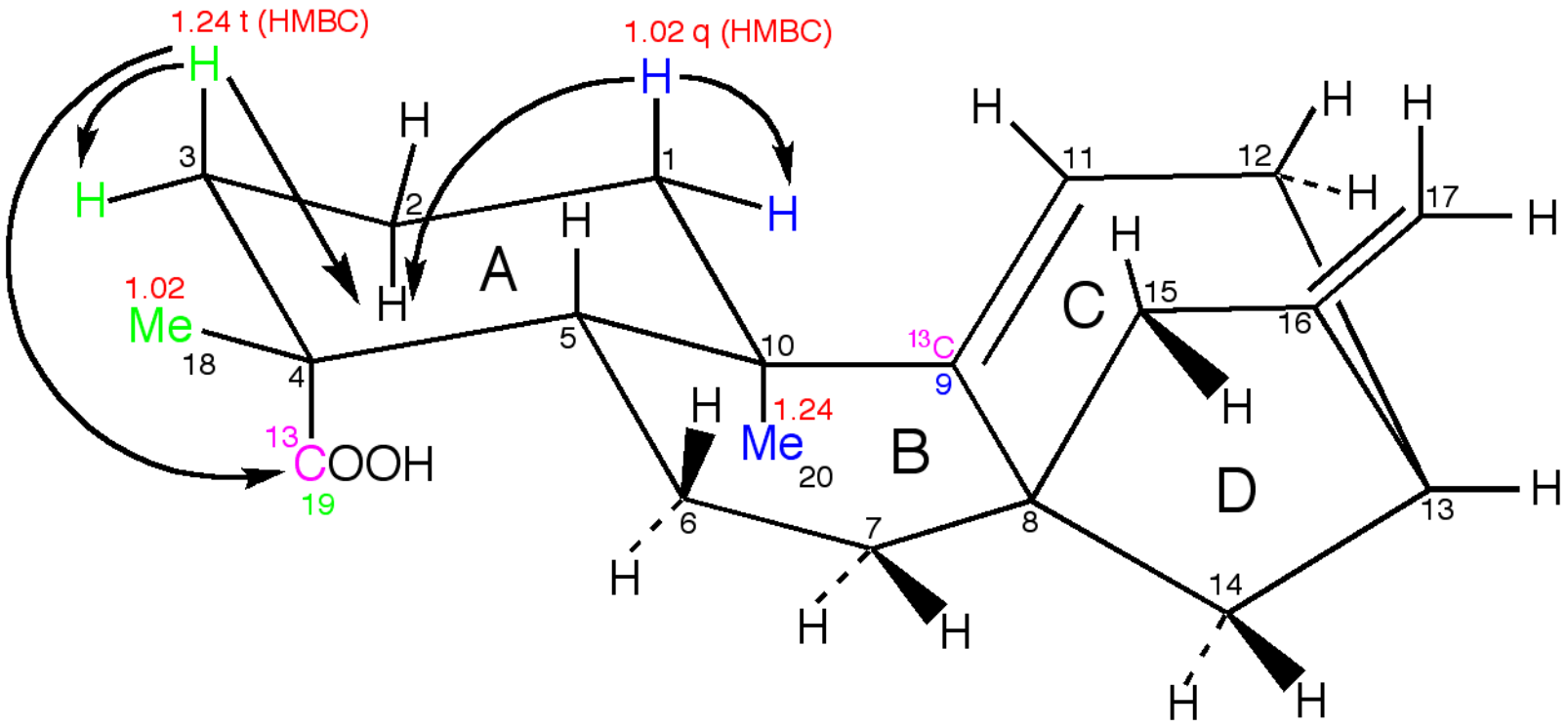
Correct Assignments: $d_N(^{13}C) = 1.17$

Incorrect Assignments: $d_N(^{13}C) = 1.26$

KA: Correct Assignment



KA: Incorrect Assignment



KA: HMBC Traces