RUNNING DR-Thermo

Reactions:

The reaction set (representing the thermodynamic constraints) should be saved in reactions.txt file in the 'reactions' folder. Each reaction in this file is identified by the Reaction ID (RID) and each compound in the reaction is identified by the Compound ID (CID).

Data

The corresponding thermodynamic data for standard transformed reaction Gibbs free energies (dG'_r) and formation Gibbs free energies (dG'_f) corresponding to each RID and CID should be saved in the 'reaction_dG.txt' and 'formation_dG.txt' files respectively. When there are multiple Gibbs free energy measurements for the same reaction (or compound), either under the same or different reaction conditions, they should be included in the data with the same RID (or CID). The presence of such replicates would increase the redundancy of the DR algorithm. The pKa values should be saved as a pKa.mat file.

(An additional .txt file with the replicates of reaction Gibbs energies, as obtained from the NIST-DECR database, is included in the "Cache" folder with the name "reaction_dG 2.txt")

Cache

The group decomposition matrix should be saved as G.mat. The kegg_pkas.mat is the Kegg file with the acid dissociation constants, identified by the CIDs.

Main

Main.m is the file to be used to run DR-thermo. It generates all "chemical energy" measurements using the inverse Legendre transform and reads the reaction set to determine the constraint matrix (stoichiometric matrix S). From the available measurement data, it then determines the missing reaction and formation Gibbs energies. Thereafter, all available measurements are reconciled and the unavailable measurements are categorized as either observable or unobservable. For the unobservable variables, the group contribution estimates are generated using all available data and imputed onto the estimation step in the coaptation problem. Thereafter, all the reconciled data, in the form of standard Gibbs energies of reaction and formation corresponding to each reaction and compound in the RIDs and CIDs arrays, is compiled and stored in a "result" structure array. The RIDs and the CIDs corresponding the unobservable reactions and compound Gibbs energies are also displayed. When an unobservable variable cannot be estimated by group contributions, its resulting value in the result structure array will be represented by "NaN". The RIDs and CIDs corresponding to these unobservable variables are stored in the RIDS_GC_NA and CIDS_GC_NA arrays. If these arrays are not present in the result, it means that group contribution estimates have been imputed for all unobservable variables.

Use

The algorithm can be used to curate experimental thermodynamic data from different sources to generate reliable data-banks such that all measurements are modified to be thermodynamically consistent and any missing reaction or formation Gibbs energies can be predicted either directly or using group contributions. The algorithm can also be used to identify "unobservable" reaction and formation Gibbs energies, which can be targets for future experiments.