

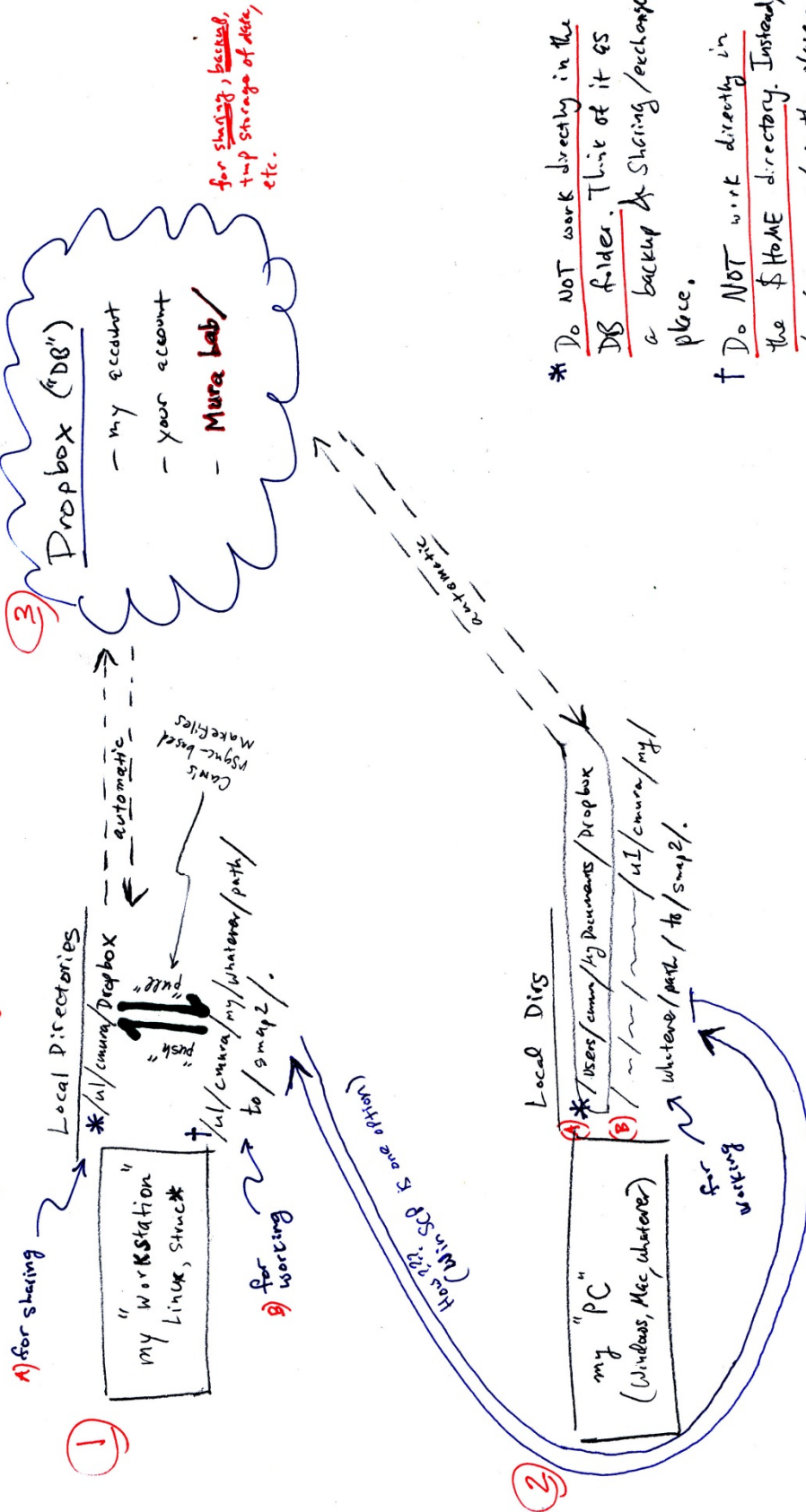
Cam's Unix Notes, 2 — **Computing locales in our StrucBio system**; Use Dropbox, rsync, and WinSCP to build yourself a robust workflow

When working across several platforms (Windows, Mac, Linux/Unix) and on multiple computers, data management, research activities (*e.g.*, processing diffraction datasets), manuscript preparation, and many other tasks can become quite challenging. Smooth integration is extremely difficult to achieve. Dropbox ('DB') helps to some extent, but it is not a cure-all: Because of file-locking and other issues, experience shows that DB should be viewed less as a dynamic workspace (where one actively *does* work, *edits* files, *etc.*), and instead more as a utility for two important purposes – (*i*) **sharing & exchanging files** and other collaborative data between multiple individuals, and (*ii*) keeping **automatic backups** of valuable files (not every single piece of data, otherwise the DB space quota will be exceeded). In other words, for the purposes described herein, DB is used in a more 'passive' rather than 'active' manner.

To see how DB can play a role in your personal workflow, consider the three key locales in our StrucBio system: (1) my struc* workstation [Linux], (2) my PC laptop [Windows, Mac], and (3) my Dropbox account. These *computing locales* are numbered in the diagram below (pg 2), which illustrates a system (a *workflow*) that I've evolved over the past few years and found to be highly robust and effective. To understand this system, study the diagram and ask me questions. Also, it is important to note that locales (1) and (2) each contain an important pair of *local* directories – (A) the DB directory itself (specified when you installed and configured DB on your machine) and (B) a "working directory" space, where you *do* the actual work (write a paper, make a PyMOL figure, *etc.*). On our struc Linux workstations (*locale-1*), this *working directory* might be the path /u1/cmura/Dropbox/, and on my personal desktop or laptop (*locale-2*) this working area might be something like C:\Users\cmura\Documents\u1\cmura\jove_dna_rna_prot. The two crucial – and perhaps least obvious – steps that make this system work are (*i*) being able to flexibly sync (in 'push'/'pull' directions) between the two directories (A) and (B) in *locale-1* [Linux], achieved via the Unix 'rsync' command^[1], and (*ii*) being able to flexibly sync (in 'push'/'pull' modes) between the working directory (B) in *locale-1* [Linux] and the working directory (B) in *locale-2* [Windows/Mac], which can be easily achieved on Windows via the free program WinSCP. Other possible workflow solutions that I've explored, but which are beyond the scope of this piece, integrate Samba/CIFS technology to facilitate interoperability between *locales 1* (Linux) and *2* (Windows/Mac); some readers may find such approaches useful.

[1] These rsync commands can become quite complex (*e.g.*, intricate exclusion and inclusion rules can be built-up); therefore, I have found it useful to wrap these commands into a Makefile so that one can simply type 'make push_real', 'make pull_fake', *etc.*

Three key computing locales:



- * Do NOT work directly in the DB folder. Think of it as a backup & sharing/exchange place.
- † Do NOT work directly in the \$HOME directory. Instead, /u1/USERNAME/ is the place for that, on LINUX/structbio workstations.
- Do NOT store your local DB folder in \$HOME on StructBio system.