Identification of new antibacterial targets in RNA polymerase of *Mycobacterium tuberculosis* by detecting positive selection sites.

QingBiao Wang[a,1], Yiqin Xu [a,1] Zhuoya Gu [a,1], Nian Liu[b], Ke Jin[a], Yao Li[a],

M. James C. Crabbe[c,d] ,Yang Zhong[e,a*]

[a] School of Life Sciences, Fudan University, Shanghai 200433, China
[b] Fudan International School, High School Affiliated to Fudan University, Shanghai 200433, China
[c] Wolfson College, Oxford University, Oxford, OX2 6UD, U.K.
[d] Institute of Biomedical and Environmental Science and Technology, University of Bedfordshire, Luton, LU1 3JU, U.K.
[e] Institute of Biodiversity Science and Geobiology, Tibet University, Lhasa 850000, China

[*] Corresponding author at: Institute of Biodiversity and Geobiology, Tibet University, Lhasa 850000, China
E-mail address: yangzhong@fudan.edu.cn (Y. Zhong).
Tel.: +86 21 55664436; fax: +86 21 65642263.
[1] These authors contributed equally to this work.

Abstract:

Bacterial RNA polymerase (RNAP) is an effective target for antibacterial treatment. In order to search new potential targets in RNAP of *Mycobacterium*, we detected adaptive selections of RNAP related genes in 13 strains of *Mycobacterium* by phylogenetic analysis. We first collected sequences of 17 genes including rpoA, rpoB, rpoC, rpoZ, and sigma factor A-M. Then maximum likelihood trees were constructed, followed by positive selection detection. We found that sigG shows positive selection along the clade (*M. tuberculosis*, *M. bovis*), suggesting its important evolutionary role and its potential to be a new antibacterial target. Moreover, the regions near 933Cys and 935His on the rpoB subunit of *M. tuberculosis* showed significant positive selection, which could also be a new attractive target for anti-tuberculosis drugs.

Key words: phylogenetic analysis, selective pressure, target sites, RNA polymerase, Mycobacterium tuberculosis

1. Introduction

The bacterial RNA polymerase (RNAP）subunit sequences are highly conserved and play an important role in broad-spectrum antibacterial targets searching (Chopra, 2007; Darst, 2004). RNAP is also a key target for the therapy against *Mycobacterium* especially for *Mycobacterium tuberculosis*. The RNAP enzyme comprises a multi-subunit αββ'ω core which forms a 'crab-claw-like' structure, and the β and β'-subunits form the main components of each 'pincer' of the claw. The commonly used anti-tuberculosis (anti-TB) drug rifampicin binds to the β-subunit of the RNAP encoded by rpoB and inhibits transcription.

However, multidrug-resistant (MDR) strains of *Mycobacterium tuberculosis* are resistant to some anti-TB drugs targeting RNAP including rifampicin and isoniazid (Gandhi et al., 2010). Resistance to rifampicin can result from several reasons; first, resistance to rifampicin typically contains substitution of residues in or close to this drug binding site on bacterial RNAP (Campbell et al., 2001). More than 95% of *Mycobacterium tuberculosis* clinical strains resistant to rifampicin have at least one mutation in an 81-basepair region of rpoB called rifampicin resistance determining region (RRDR) (Ramaswamy and Musser, 1998). A second reason for the emergence of rifampicin resistance is the avoidance of fitness deficit in MDR strains. Although experimental work has shown that drug resistance in bacteria was often associated with a fitness deficit (Andersson and Hughes, 2010; Andersson and Levin, 1999), this fitness cost of MDR *Mycobacterium tuberculosis* strains can be mitigated by compensatory mutations on rpoA or rpoC, which causes the mutant strains to thrive successfully (Comas et al., 2012). Therefore, there is a stringent need for new antibacterial agents targeting at bacterial RNAP but distinct from the rifampicin binding site.

A common method to search for key sites in proteins is to confirm the function by site-directed mutagenesis, which is time-consuming and labor-intensive. However, molecular evolutionary studies provide another opportunity. According to molecular evolution theories, the sequence of a gene with a key function should be highly conserved among species. As a result, mutations at conserved sites usually lead to deleterious effects on its function, and mutants are eliminated by negative selection. Only the mutations benefitting the species' survival and reproduction could be fixed by positive selection, suggesting that the sites undergoing frequent positive or negative selections are crucial to the gene function. Therefore, it is possible to find the ideal target site of anti-TB drugs by detecting the sites undergoing adaptive selections in the RNA polymerase.

In this study, we constructed phylogenetic trees with 17 corresponding genes separately which have great influence on the function of the RNAP enzyme in 13 *Mycobacterium* species. Based on these phylogenetic trees, we detected selective pressure on each branch and searched for the positive selection sites. Moreover, we predicted the tertiary structure of RNAP using homology modeling, and marked potential antibacterial therapeutic target sites. Finally, we compared positive selection sites with the molecular docking results of the RNAP enzyme in *Mycobacterium tuberculosis* and broad-spectrum antibacterial drugs. In short, this study provides a novel way to seek for drug targets and develop new anti-TB drugs.

2. Materials and methods

2.1 Collection and Identification of the related genes to RNAP enzyme

In this study, we chose 17 corresponding genes which have significant impacts on function of the RNAP enzyme. These genes were rpoA, rpoB, rpoC, rpoZ and 13 sigma factors from A to M. We collected the nucleotide sequences of these genes in 13 strains of *Mycobacterium* genus and 2 strains which are *Gordonia bronchialis* DSM 43247 and *Rhodocshaoccus opacus* B4 as outgroups in NCBI (http://www.ncbi.nlm.nih.gov/) for the phylogenetic analysis (Table 1). Based on the 17 diverse genes of *Mycobacterium tuberculosis H37Rv*, the relevant orthologous genes in other 12 strains were obtained by best reciprocal hit in BLAST (Bork et al., 1998; Tatusov et al., 1997).

2.2 Construction of phylogenetic trees in each relevant genes

Each homologous gene nucleotide sequence was first translated into its amino acid sequence and aligned by the Emboss software package (Rice et al., 2000). Then we realigned nucleotide sequences of each gene based on the result of amino acid sequence alignments.

The best Akaike information criterion (AIC) model of every genes in all strains was obtained with PAUP 4.0 (Swofford, 1993) and Modeltest 3.7 (Posada and Crandall, 1998). Based on the AIC model, the maximum likelihood (ML) trees of each gene were constructed using PHYML 3.0 (Guindon et al., 2010) with 1000 bootstrap replications.

2.3 Detection of selective pressures

The analysis of adaptive evolution was performed with the CODEML algorithm in PAML 4.4 (Yang, 2007). We used the topology of maximum likelihood tree and the result of codon alignment of corresponding genes to calculate the selective pressure on each branch. Two different approaches were used, branch models and branch-site models. For both approaches, branches in the tree were divided a priori into foreground and background categories, and only foreground lineages may have experienced positive selection. The branches and sites under positive selection were identified by likelihood ratio tests (LRTs) which were applied by comparing the differences in log-likelihood (In L) values between two nested models, which are single ratio and free ratio using a $\chi 2$ distribution, and the differences in the number of parameters between two nested models is the degrees of freedom.

3. Results
3.1 Phylogenetic analysis of the related genes to RNAP enzyme

Sigma factors are components of the RNAP complex that are responsible for binding to the RNAP complex, promoter recognition and separating DNA strands (Paget and Helmann, 2003). The number of sigma factors is extremely diverse between bacteria, ranging from 3 in *Helicobacter pylori* to 63 in *Streptomyces coelicolor.* In our study the relevant orthologous genes, except for rpoA, rpoB, rpoC and rpoZ, only four sigma factors including sigA, sigB, sigD and sigE were found in the genome of all 13 strains on *Mycobacterium* genus and 2 outgroup strains. It was interesting that we only found the orthologous gene of sigC in 7 pathogenic *Mycobacterium* species, including *M. abscessus*, *M. avium*, *M. bovis*, *M. leprae*, *M. marinum*, *M. tuberculosis* and *M. ulcerans*. This was in line with the viewpoint that sigC governs expression of an important pathogenic *Mycobacterium* regulon and is essential for lethality in host but not required for bacterial survival (Sun et al., 2004). As a typical intracellular parasite, *M. leprae* has the smallest genome in *Mycobacterium* due to gene elimination. It is worth mentioning that *M. leprae* only contains rpoA, rpoB, rpoC, rpoZ, sigA, sigB, sigC and sigE genes while the other sigma factors were either pseudogenes or lost. Because every sigma factor has its own specificity which allows the initiation of transcription of different subsets of genes (Manganelli et al., 2004), we inferred that sigma factors which are most relevant to virulence are sigA, sigB, sigC and sigE.

The phylogenetic trees based on different related genes to the RNAP enzyme had similar architecture. Six out of seven pathogenic *Mycobacterium* strains were clustered in one big clade, the remaining one, *M. abscessus* strain, was on the basal branch of *Mycobacterium* genus. In addition, the 5 nonpathogenic *Mycobacterium* strains including *M. gilvum*, *M. vanbaalenii*, *M. sp. JLS*, *M. sp.*

KMS and *M. sp. MCS* that are considered as decomposers of aromatic compounds in the soil were clustered in a clade between the 6 pathogenic *Mycobacterium* strains and the *M. abscessus* strain (see Figure 1). Because *Mycobacterium abscessus* is the most resistant and rapidly growing *Mycobacterium* and causes a wide range of clinical infectious diseases (Huang et al., 2010), we inferred that pathogenesis-related protein domains have greatly changed in *Mycobacterium abscessus*, therefore further study needs more attention on the virulence factors of this strain and corresponding target drugs. In addition, *Mycobacterium smegmatis* as a saprophytic *Mycobacterium* has none of the pathogenic properties described for *Mycobacterium tuberculosis*. But this strain is a potential cause of inflammation of soft tissues (Reyrat and Kahn, 2001; Shimizu et al., 2012). Therefore *Mycobacterium smegmatis* was not always clustered in the clade of nonpathogenic *Mycobacterium*, and sometimes it was on the basal branch of the *Mycobacterium* genus, indicating that a great change has occurred in its gene sequence and relevant protein structure.

3.2 Identification of positive selection in adaptive evolution.

We found that no positive selection was detected in the phylogenetic trees of corresponding genes to the RNAP enzyme including rpoA, rpoZ, sigA, sigC, sigD, sigE, sigH, sigI, sigK, sigL and sigM by using the CODEML algorithm in PAML 4.4.

In the phylogenetic tree of rpoB gene (Figure 1A), we found positive selection using branch model on the branch including ((*M. tuberculosis, M. bovis*), *M. leprae*) and (*M. tuberculosis, M. bovis*), and there were 3 positively selected sites which are 656His (A→H), 933Cys (V→C), and 935His (K→H) detected on the branch (*M. tuberculosis, M. bovis*) using the branch-site model (Table 2). For the rpoC gene (Figure 1B), positive selections were found on the branch including (*M.marinum, M. ulcerans*), *M. leprae* and *M. avium*, and 1 positively selected site 952Cys (L→C) was detected on the branch (*M. marinum, M. ulcerans*) with the branch-site model (Table 2).

In this study, we also found positive selection on sigma factors consisting of sigB, sigF, sigG and sigJ. In the phylogenetic tree of the sigB gene (Figure 2A), we found positive selection on the branch (((*M. marinum, M. ulcerans*), ((*M. tuberculosis, M. bovis*) *M. leprae*)), *M. avium*), but no positively selected sites were detected (Table2). For sigF gene (Figure 2B) positive selection was found on the branch (((*M. marinum, M. ulcerans*), (*M. tuberculosis, M. bovis*)), *M. avium*), but only 1 positively selected site 14Asn (S→N) was detected (Table 2). For sigG gene (Figure 2C), we found positive selection on the branch (((*M. marinum, M. ulcerans*), (*M. tuberculosis, M. bovis*)), *M. avium*), (*M. tuberculosis, M. bovis*) and *M. avium*, in addition, we found 3 positively selected sites 59Thr(L→T), 91Leu Y→L), and 316Leu(T→L) on the branch ((*M. marinum, M. ulcerans*), (*M. tuberculosis, M. bovis*)), *M. avium*), 2 positively selected sites 174Asn(S→N) and 306Tyr(E→Y) on the branch (*M. tuberculosis, M. bovis*) and 2 positively selected sites 204Gly(Q→G), 206Arg(K→R) on the branch *M. avium*(Table 2). For sigJ gene (Figure 2D), we found positive selection on 2 branches, *M. avium* and *M. marinum*, we also found 3 positively selected sites including 3Gly(F→G), 8Glu(P→E), and 9Val(H→V) on the branch *M. avium* and only 1 positively selected site 198Arg(K→R) on the branch *M. marinum* (Table 2).

4. Discussion

In this study, we found many positively selected sites in components of the RNAP complex enzyme in the *Mycobacterium* genus including rpoB, rpoC and three sigma factors, sigF, sigG and sigJ. Moreover, the positively selected sites of sigG were much more than the other two sigma

factors. In addition, positively selected sites on the branch (*M. tuberculosis, M. bovis*) only appeared in sigG, indicating that sigG may be an important potential antibacterial therapeutic target subunit. Furthermore, sigG would be significantly upregulated by *Mycobacterium tuberculosis* in human macrophage infection (Cappelli et al., 2006; Lee et al., 2008). Therefore, it can be inferred that sigG plays an important role in the evolution of *M. tuberculosis* and *M. bovis*. It appears that sequence alignments of the sigma[70] family members have four conserved regions that can be further divided into sub-regions (Lonetto et al., 1992). It is interesting that most positively selected sites of sigF, sigG and sigJ were concentrated on region 2 and region 4 which banded on the upstream of the promoter and participate in DNA strand opening (Paget and Helmann, 2003). As a result, it is necessary to detect whether there were mutants on the upstream sequences of corresponding regulation genes.

It is well-known that rpoB and rpoC are two important components, β and β'-subunits, of a multi-subunit αββ'ω core which forms a "crab claw-like" structure. As two "pincers" of the claw, they form a groove that accommodates the template DNA and provides a catalytic site for phosphodiester bond formation, a secondary channel for incoming nucleotides, and a separate exit for the growing RNA transcript (Borukhov and Nudler, 2008; Koch et al., 2014). Many medical inhibitors including rifampicin, tagetitoxin and myxopyronin specifically target rpoB and rpoC subunits of the RNAP enzyme (Artsimovitch et al., 2011; Malinen et al., 2012; Mukhopadhyay et al., 2008). According to the docking results of *Mycobacterium tuberculosis* RNAP and rifampicin (Lin and Mandal et al.,2017), RNAP and rifampicin interactions are similar to those in previously reported structures of *Thermus aquaticus* RNAP, and *Thermus Thermophilus* RNAP with rifampicin (Campbell and Korzheva et al., 2001; Artsimovitch and Vassylyeva et al., 2005). The binding sites are concentrated entirely among R173 S437 L436 Q438 S456 G459 Q435 L458 V176 R465 N493 I497 P489 R613 R454 D441 H451 H680 F439 on rpoB subunit. Beyond the rpoB subunit, the rpoC subunit is also a good target to create a new drug to inhibit *Mycobacterium tuberculosis* RNAP; there is a new small molecule inhibitor of *Mycobacterium tuberculosis* RNAP named Nα-aroyl-N-aryl-phe-nylalaninamides (AAPs) binding on both rpoB and rpoC subunits. The interaction region of D-AAP1 (the prototype of AAPs) and RNAP contains V568 R562 G566 Y480 P477 V475 on rpoB subunit and F850 H854 P827 N826 I851 I832 E848 R834 L847 on rpoC subunit. We compared our positively selected sites on rpoB and rpoC with the binding sites of rifampicin and D-AAP1 (PDB code:5UHG) on *Mycobacterium tuberculosis* RNAP protein with PYMOL (Figure 3). The three positivel selection sites, H656, C933 and H935 on rpoB encompass the center of the interaction region of rifampicin on *Mycobacterium tuberculosis* RNAP protein. Based on the location of the three positively selection sites, the H656 on rpoB and C952 on rpoC may cause surface conformational changes in other *Mycobacterium* genus, while the C933 and H935 are close to each other and in the vicinity of similar binding sites to tagetitoxin in bacterial RNAP from *T. thermophilus* (Vassylyev and Svetlov et al.,2005), but separated by a helix of a dozen amino acids; the region near them might be a new potential therapeutic target.

Furthermore, both two positively selected sites C933 and H935 on rpoB are close to the "switch region" which is located at the base of clamp also called the "pincer" and serves as the hinge on which the clamp swings in clamp opening and closure (Cramer, 2002; Cramer et al., 2001; Gnatt et al., 2001). The switch region adopts different conformations in open and closed clamp conformational states. Moreover, the clamp must to be open to permit DNA to enter the active-center cleft during early stage of transcription initiation and is required to close to keep

DNA in the active-center cleft during later stages of transcription initiation and elongation. Based on the similarity of the location of our positively selected sites on rpoB and binding sites to tagetitoxin on bacterial RNAP, we suggest RNAP switch region to be an attractive target for new antibacterial therapeutic agents for *Mycobacterium tuberculosis*. Firstly, the comprised residues of the switch region were highly conserved in all species of *Mycobacterium* genus even in Gram-positive bacterial RNAP and Gram-negative bacterial RNAP, providing a foundation for broad-spectrum activity of drugs. Secondly, two positive selection sites were found close to the "switch region" of RNAP in *Mycobacterium tuberculosis*, indicating the specific functional structure which could be a specific drug target. Thirdly, this switch region is far away from the binding sites for typical inhibitors of bacterial RNAP and also not in the rifampicin resistance-determining region (RRDR) (Chopra, 2007). The large distance therefore provides a basis for absence of cross-resistance with other classic inhibitors of RNAP, especially rifampicin. Moreover, the distance ensures a low probability of occurrence of putative compensatory mutations which decreases fitness cost (Comas et al., 2012). Finally, the ligand binding sites in this region contain a nearly completely enclosed hydrophobic pocket which offer a basis for efficient "druggability" by multiple chemotypes (Mukhopadhyay et al., 2008), just like HIV-1 reverse transcriptase NNRTI sites which have high druggability and utility (Sluis-Cremer et al., 2004).

Overall our study provides a new way to search for potential antibacterial therapeutic targets in the RNA polymerase of *Mycobacterium tuberculosis* by detecting positive selection sites. In 13 strains on *Mycobacterium* genus, new potential targets possessing the highest fitness costs of resistance and lowest likelihood of compensation have been detected by phylogenetic analysis and selection pressure detection. Fortunately as these findings are not in the known binding region, we can regard these positive selection sites as potential inhibitor targets.

Authors' contributions

ZG, YX, QW, NL and YZ conceived the project, performed the calculations and analyzed the results. All authors wrote the paper, and read and approved the final manuscript.

Funding

Competing interests

The authors have declared that no competing interests exist.

References

Andersson DI and Hughes D (2010) Antibiotic resistance and its cost: is it possible to reverse resistance? *Nature reviews Microbiology***8**:260-271.

Andersson DI and Levin BR (1999) The biological cost of antibiotic resistance. *Current Opinion in Microbiology***2**:489-493.

Artsimovitch I, Svetlov V, Nemetski SM, Epshtein V, Cardozo T and Nudler E (2011) Tagetitoxin

inhibits RNA polymerase through trapping of the trigger loop. *The Journal of biological chemistry***286**:40395-40400.

Bork P, Dandekar T, Diaz-Lazcoz Y, Eisenhaber F, Huynen M and Yuan Y (1998) Predicting function: from genes to genomes and back. *Journal of molecular biology***283**:707-725.

Borukhov S and Nudler E (2008) RNA polymerase: the vehicle of transcription. *Trends in Microbiology***16**:126-134.

Campbell EA, Korzheva N, Mustaev A, Murakami K, Nair S, Goldfarb A and Darst SA (2001) Structural Mechanism for Rifampicin Inhibition of Bacterial RNA Polymerase. *Cell***104**:901-912.

Campbell, E. A. and N. Korzheva, et al. (2001). "Structural Mechanism for Rifampicin Inhibition of Bacterial RNA Polymerase." Cell **104** (6): 901-912.

Cappelli G, Volpe E, Grassi M, Liseo B, Colizzi V and Mariani F (2006) Profiling of Mycobacterium tuberculosis gene expression during human macrophage infection: Upregulation of the alternative sigma factor G, a group of transcriptional regulators, and proteins with unknown function. *Research in Microbiology***157**:445-455.

Chopra I (2007) Bacterial RNA polymerase: a promising target for the discovery of new antimicrobial agents. *Current opinion in investigational drugs (London, England : 2000)***8**:600-607.

Comas I, Borrell S, Roetzer A, Rose G, Malla B, Kato-Maeda M, Galagan J, Niemann S and Gagneux S (2012) Whole-genome sequencing of rifampicin-resistant Mycobacterium tuberculosis strains identifies compensatory mutations in RNA polymerase genes. *Nat Genet***44**:106-110.

Cramer P (2002) Multisubunit RNA polymerases. *Current opinion in structural biology***12**:89-97.

Cramer P, Bushnell DA and Kornberg RD (2001) Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution. *Science (New York, NY)***292**:1863-1876.

Darst SA (2004) New inhibitors targeting bacterial RNA polymerase. *Trends in biochemical sciences***29**:159-160.

Gandhi NR, Nunn P, Dheda K, Schaaf HS, Zignol M, van Soolingen D, Jensen P and Bayona J (2010) Multidrug-resistant and extensively drug-resistant tuberculosis: a threat to global control of tuberculosis. *The Lancet***375**:1830-1843.

Gnatt AL, Cramer P, Fu J, Bushnell DA and Kornberg RD (2001) Structural basis of transcription: an RNA polymerase II elongation complex at 3.3 A resolution. *Science (New York, NY)***292**:1876-1882.

Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W and Gascuel O (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic biology***59**:307-321.

Huang YC, Liu MF, Shen GH, Lin CF, Kao CC, Liu PY and Shi ZY (2010) Clinical outcome of Mycobacterium abscessus infection and antimicrobial susceptibility testing. *Journal of microbiology, immunology, and infection = Wei mian yu gan ran za zhi***43**:401-406.

Koch A, Mizrahi V and Warner DF (2014) The impact of drug resistance on Mycobacterium tuberculosis physiology: what can we learn from rifampicin? *Emerg Microbes Infect***3**:e17.

Lee JH, Geiman DE and Bishai WR (2008) Role of stress response sigma factor SigG in Mycobacterium tuberculosis. *Journal of Bacteriology***190**:1128-1133.

Lin W, Mandal S, Degen D, et al. Structural Basis of Mycobacterium tuberculosis Transcription and Transcription Inhibition.[J]. Molecular Cell, 2017, 66(2):169.

Lonetto M, Gribskov M and Gross CA (1992) The sigma 70 family: sequence conservation and evolutionary relationships. *Journal of Bacteriology***174**:3843-3849.

Malinen AM, Turtola M, Parthiban M, Vainonen L, Johnson MS and Belogurov GA (2012) Active site opening and closure control translocation of multisubunit RNA polymerase. *Nucleic acids research***40**:7442-7451.

Manganelli R, Proveddi R, Rodrigue S, Beaucher J, Gaudreau L and Smith I (2004) σ Factors and Global Gene Regulation in Mycobacterium tuberculosis. *Journal of Bacteriology***186**:895-902.

Mukhopadhyay J, Das K, Ismail S, Koppstein D, Jang M, Hudson B, Sarafianos S, Tuske S, Patel J, Jansen R, Irschik H, Arnold E and Ebright RH (2008) The RNA polymerase "switch region" is a target for inhibitors. *Cell***135**:295-307.

Paget MSB and Helmann JD (2003) Protein family review - The sigma(70) family of sigma factors. *Genome Biology***4**.

Posada D and Crandall KA (1998) MODELTEST: testing the model of DNA substitution. *Bioinformatics (Oxford, England)***14**:817-818.

Ramaswamy S and Musser JM (1998) Molecular genetic basis of antimicrobial agent resistance inMycobacterium tuberculosis: 1998 update. *Tubercle and Lung Disease***79**:3-29.

Reyrat JM and Kahn D (2001) Mycobacterium smegmatis: an absurd model for tuberculosis? *Trends Microbiol***9**:472-474.

Rice P, Longden I and Bleasby A (2000) EMBOSS: the European Molecular Biology Open Software Suite. *Trends in genetics : TIG***16**:276-277.

Shimizu F, Hatano Y, Okamoto O, Katagiri K, Fujiwara S, Sato S, Kato A, Uezato H, Asato Y and Takahashi K (2012) Mycobacterium smegmatis soft tissue infection. *International journal of dermatology***51**:1518-1520.

Sluis-Cremer N, Temiz NA and Bahar I (2004) Conformational changes in HIV-1 reverse transcriptase induced by nonnucleoside reverse transcriptase inhibitor binding. *Current HIV research***2**:323-332.

Sun RG, Converse PJ, Ko C, Tyagi S, Morrison NE and Bishai WR (2004) Mycobacterium tuberculosis ECF sigma factor sigC is required for lethality in mice and for the conditional expression of a defined gene set. *Molecular Microbiology***52**:25-38.

Swofford DL (1993) Phylogenetic analysis using parsimony. *Champaign, IL: Illinois Natural History Survey*.

Tatusov RL, Koonin EV and Lipman DJ (1997) A genomic perspective on protein families. *Science (New York, NY)***278**:631-637.

Vassylyev D G, Svetlov V, Vassylyeva M N, et al. Structural basis for transcription inhibition by tagetitoxin[J]. Nature Structural & Molecular Biology, 2005, 12(12):1086-1093.

Yang Z (2007) PAML 4: phylogenetic analysis by maximum likelihood. *Molecular biology and evolution***24**:1586-1591.