Supporting Information

Quantitative characterization of local protein solvation to predict solvent effects on protein structure

Vincent Vagenende[1]* and Bernhardt L. Trout[2]

(1) Bioprocessing Technology Institute, A*STAR (Agency for Science, Technology and Research), 20 Biopolis Way #06-01 Centros, Singapore 138668

(2) Department of Chemical Engineering, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, E19-502b, Massachusetts 02139, USA

*Corresponding author: Vincent_Vagenende@bti.a-star.edu.sg

Supporting Tables

Table S1: Simulation systems and global protein solvation properties.

Protein	Constrain[a]	Time[b] [μs]	n_W [-]	n_X [-]	n_{WP}(r<5Å) [-]	n_{XP}(r<5Å) [-]	Γ_{XP} [-]
Fv D1.3	backbone	1.3	6861	725	1263.8 ± 1.6	124.5 ± 0.2	-11.1 ± 0.4
HEL	backbone	1.3	4862	514	839.2 ± 1.1	84.5 ± 0.6	-5.2 ± 0.8
HEL	all	0.7	4862	514	820.1 ± 1.2	81.8 ± 0.3	-5.9 ± 0.5

[a] Constraining of coordinates of backbone C-, N- and O-atoms or all protein atoms with a force constant of 2 kcal/(mol Å²)

[b] Total simulation time; the first 0.1 μs of the total simulation time is excluded from the analysis unless explicitly stated otherwise.

Table S2: Γ_{XP} and numbers of glycerol and water molecules at the protein surface (r < 5 Å) calculated from the first half (0.1-0.7 μs), the second half (0.7-1.3 μs) and the entire time (0.1-1.3 μs) of the simulation of HEL with constrained backbone coordinates. Note the significant increase in glycerol molecules and Γ_{XP} for the second half of the simulation.

time (μs)	Γ_{XP}	n_{WP} (r < 5 Å)	n_{XP} (r < 5 Å)
0.1-0.7	-5.9 ± 0.4	839.9 ± 1.4	84.0 ± 0.3
0.7-1.3	-4.4 ± 0.8 [a]	838.3 ± 2.1 [a]	85.0 ± 0.5 [a]
0.1-1.3	-5.2 ± 0.8	839.2 ± 1.1	84.5 ± 0.6

[a] Standard deviation estimated as the average of the four largest block times (120, 150, 200 and 300 ns).

Supporting Figures

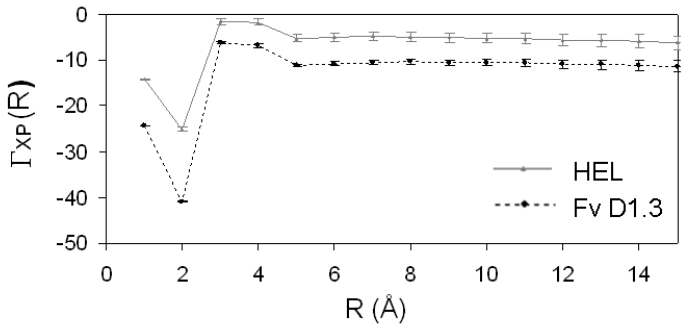

Figure S1: The global preferential interaction coefficient $\Gamma_{XP}(R)$ in function of radial distance R from the protein van der Waals surface for two proteins (HEL and D1.3).

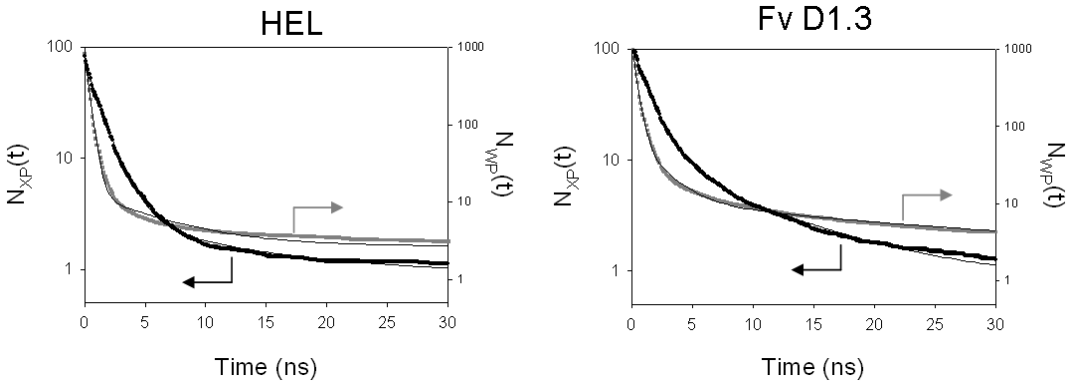

Figure S2: Survival functions $N_{XP}(t)$ (black) and $N_{WP}(t)$ (grey) and their respective fits ($N_{ZP}(t) \cong n_1 e^{-t/\tau_1} + n_2 e^{-t/\tau_2} + n_3 e^{-t/\tau_3} + c \quad with\ Z = X, W$) for HEL (left) and D1.3 (Right). Fitted functions almost entirely overlap with the respective survival functions, indicating a good quality of fit.

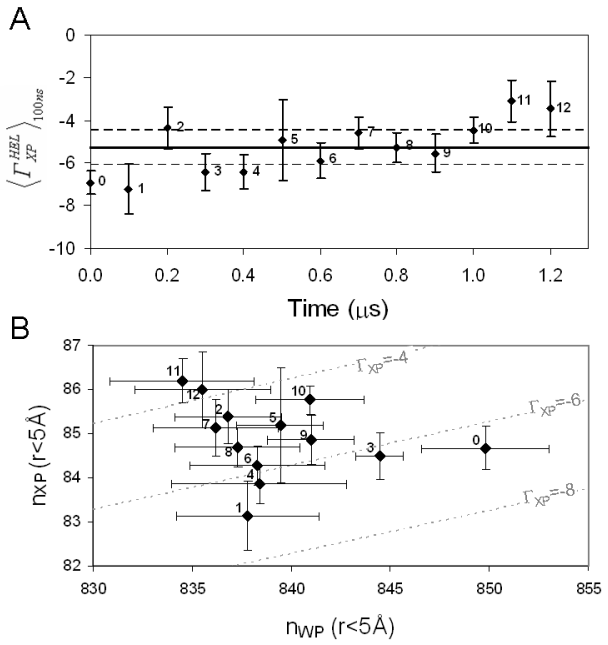

Figure S3: (A) Global preferential interaction coefficients Γ_{XP} for HEL averaged over 100 ns intervals. Horizontal lines indicate Γ_{XP} and its standard error averaged over 1.2 μs. (B) The number of water and glycerol molecules near the protein surface (r < 5 Å) averaged over subsequent 100 ns time intervals. Numbers indicate time intervals (e.g. 3 refers to 300-400 ns) and dashed lines represent iso-Γ_{XP} lines (1).

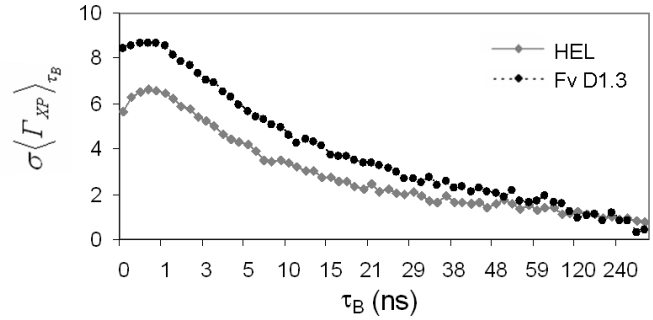

Figure S4: Standard deviations $\sigma\langle\Gamma_{XP}\rangle_{\tau_B}$ calculated for increasing block times τ_B for HEL and D1.3. Values of $\sigma\langle\Gamma_{XP}\rangle_{\tau_B}$ decrease at a slower rate for HEL than for D1.3, indicating that values of $\langle\Gamma_{XP}\rangle_{\tau_B}$ are correlated for block time averages up to 600 ns. Since standard deviation $\sigma\langle\Gamma_{XP}\rangle_{\tau_B}$ continue to decrease for increasing block time, an upper limit of the standard error $\sigma\langle\Gamma_{XP}\rangle$ can be obtained by averaging $\sigma\langle\Gamma_{XP}\rangle_{\tau_B}$ for the four largest block times, i.e. $\sigma\langle\Gamma_{XP}\rangle < \dfrac{\sigma\langle\Gamma_{XP}\rangle_{240ns} + \sigma\langle\Gamma_{XP}\rangle_{300ns} + \sigma\langle\Gamma_{XP}\rangle_{400ns} + \sigma\langle\Gamma_{XP}\rangle_{600ns}}{4}$.

.

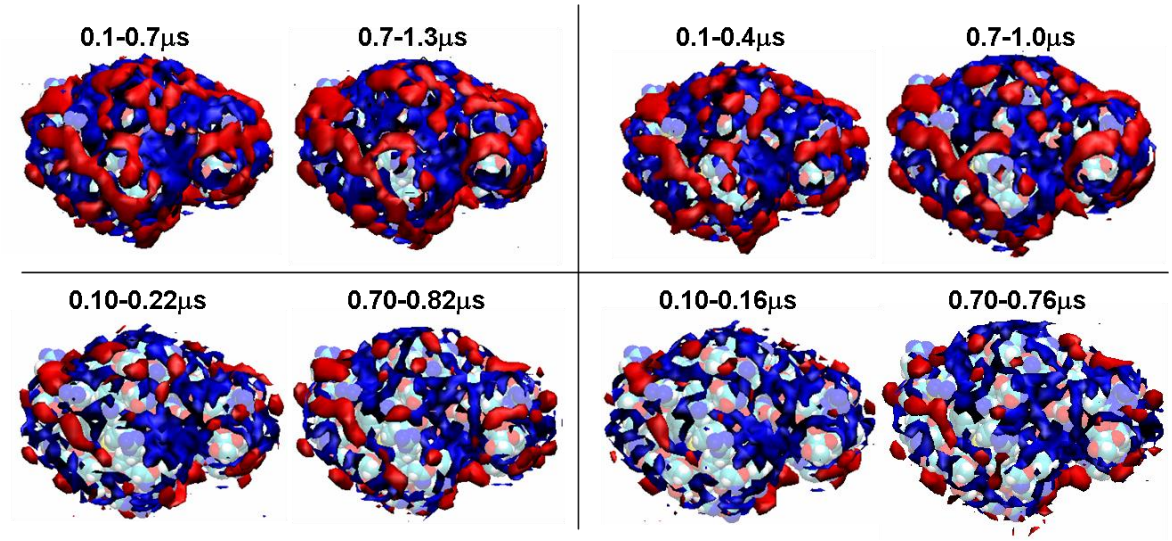

Figure S5: Local concentration maps for glycerol (red) and water (blue) at the surface of HEL averaged over the indicated time intervals. To facilitate visual assessment of convergence, maps are grouped in pairs with the same time interval length (600 ns, 300 ns, 120 ns and 60 ns, respectively). The degree of pair wise similarity and the extent of solvent regions with preferential interactions increase for longer time intervals.

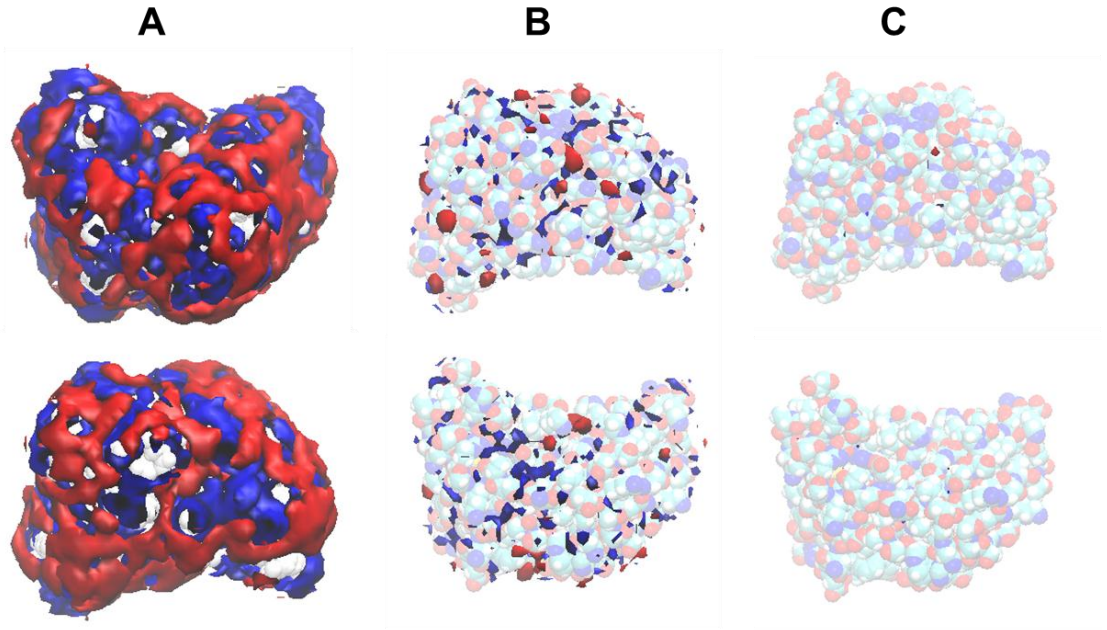

Figure S6: Local concentration maps for glycerol (red) and water (blue) at the surface of Fv D1.3 (top and bottom view) for increasing cutoff-values: $c_\alpha(\vec{r}) > c_{\alpha,bulk}$ (A), $c_\alpha(\vec{r}) > 1.5\ c_{\alpha,bulk}$ (B) and $c_\alpha(\vec{r}) > 3\ c_{\alpha,bulk}$ (C).

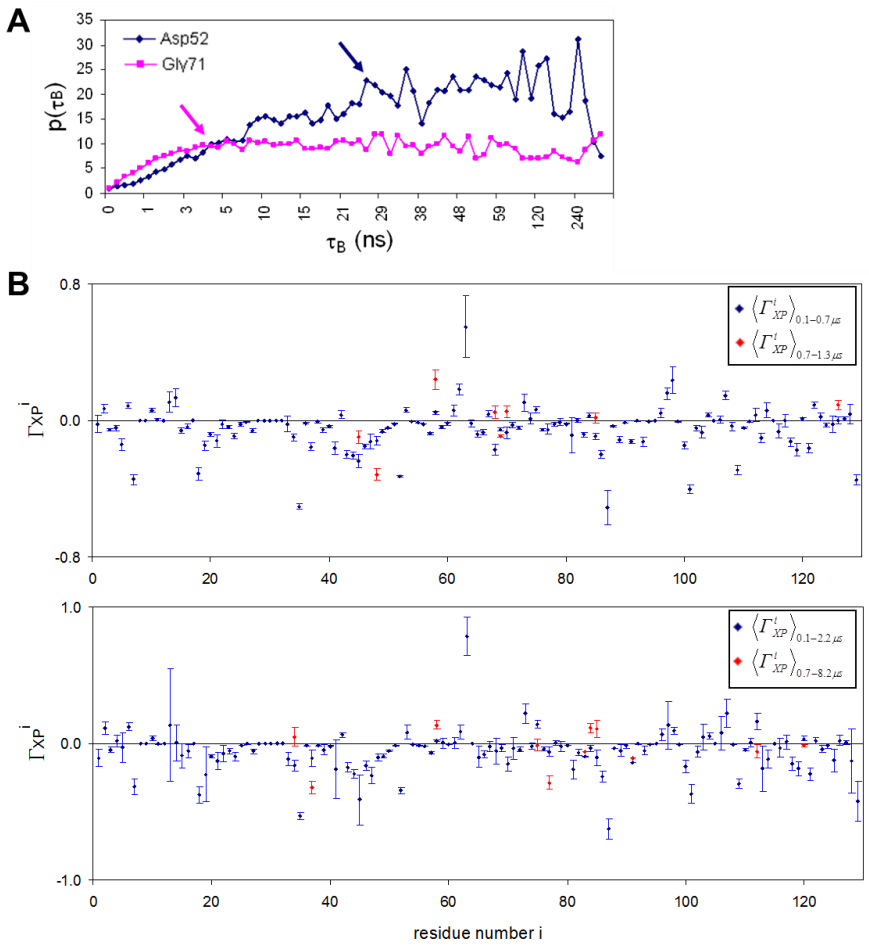

Figure S7: Statistical analysis of residue-based preferential interaction coefficients $\Gamma_{XP}^{\ i}$ of HEL: (A) $p(\tau_B)$ for $\Gamma_{XP}^{\ i}$-values of Asp52 and Gly71. Times where $p(\tau_B)$ reaches a plateau are indicated with an arrow and indicate the correlation time of $\Gamma_{XP}^{\ i}$ for the respective residues. In a similar way, $p(\tau_B)$ was plotted for each protein residue of HEL and D1.3 and we found that ~85% of solvent-accessible residues have correlation times between 5 and 100 ns. For the remaining ~15% of solvent-accessible residues, no plateau was observed indicating that convergence is slower than 100 ns.

(B) Residue-based preferential interaction coefficients $\Gamma_{XP}^{\ i}$ for HEL averaged for two distinct time intervals of 0.6 µs (Top) and 0.12 µs (Bottom). For the second interval , only those $\Gamma_{XP}^{\ i}$-values are displayed which significantly differ from $\Gamma_{XP}^{\ i}$-values in the first interval with a probability of 90%, i.e.

$$\left\{\left\langle\Gamma_{XP}^{\ i}\right\rangle_1 - 1.64\sigma\left\langle\Gamma_{XP}^{\ i}\right\rangle_1, \left\langle\Gamma_{XP}^{\ i}\right\rangle_1 + 1.64\sigma\left\langle\Gamma_{XP}^{\ i}\right\rangle_1\right\} \cap \left\{\left\langle\Gamma_{XP}^{\ i}\right\rangle_2 - 1.64\sigma\left\langle\Gamma_{XP}^{\ i}\right\rangle_2, \left\langle\Gamma_{XP}^{\ i}\right\rangle_2 + 1.64\sigma\left\langle\Gamma_{XP}^{\ i}\right\rangle_2\right\} = \emptyset$$

For $\Gamma_{XP}^{\ i}$ averaged over 0.6 µs (Top) and 0.12 µs (bottom) significant differences are found for 8 and 11 residues respectively (indicated in red). This is ~10 % of the total number of residues in HEL, which leads us to accept the statistical hypothesis that $\Gamma_{XP}^{\ i}$-values are converged within 0.12 µs. As expected the precision of almost all $\Gamma_{XP}^{\ i}$-values increases markedly as the simulation time is increased from 0.12 µs (Bottom) and 0.6 µs (Top).

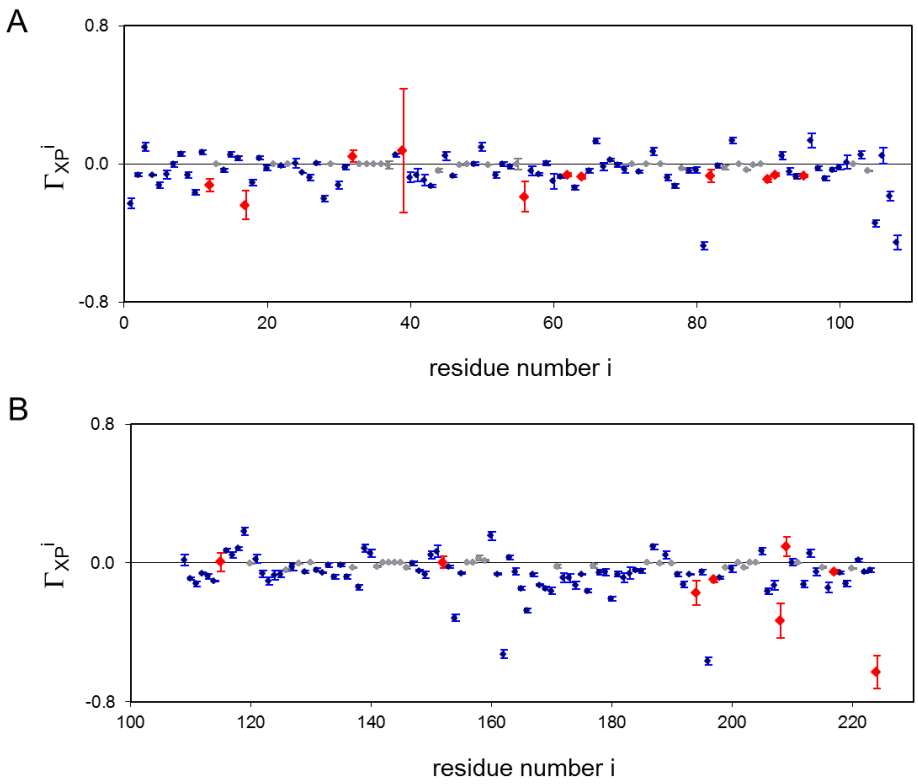

Figure S8: Residue-based preferential interaction coefficients Γ_{XP}^{i} for Fv D1.3 chain A (top) and chain B (bottom). Residues with limited solvent accessibility are shown in grey, and residues for which $p^{i}(\tau_B)$ does not maintain a plateau are colored red. All other residues are colored blue.

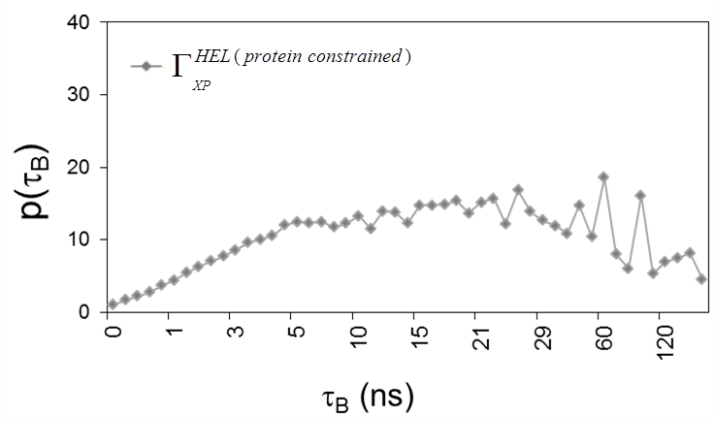

Figure S9: Statistical analysis of the global preferential interaction coefficient Γ_{XP} for the simulation of HEL with all protein coordinates are constrained. Unlike the simulation of HEL with free side-chain motion, values of $p(\tau_B)$ remain at a plateau for τ_B > 10 ns.

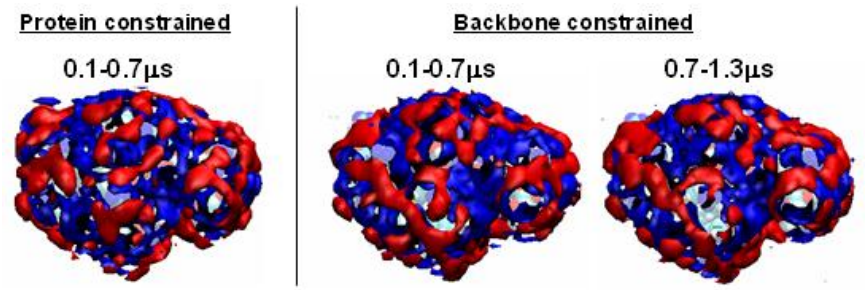

Figure S10: Effects of protein side-chain motions on local concentration maps. Local concentration maps for glycerol (red) and water (blue) at the surface of HEL averaged over the indicated time intervals. Concentration maps correspond with simulations for which all protein atom coordinates are constrained (left), or only backbone coordinates are constrained (right). The high degree of similarity between all concentration maps indicates that effects of side-chain movements on protein solvation are limited.

Supporting Movies

Movie S1: Glycerol at the protein surface of HEL near Lys 13 and Asp 18. Snapshots are rendered with VMD 1.9 for every 1 ns of the first 200 ns of the 1.3 μs simulation of HEL and include glycerol molecules with at least one atom within 4 Å of the amine group of Lys13 and within 4 Å of the carboxyl group of Asp13. Red wireframes demark solvent regions with high local glycerol concentrations ($c_X(\vec{r})$ > $1.5c_{X,bulk}$). Red wireframes demark solvent regions with high local glycerol concentrations ($c_X(\vec{r})$ > $1.5c_{X,bulk}$).

Movie S2: Glycerol in the catalytic binding pocket of HEL. Snapshots are rendered with VMD 1.9 for 10 ns steps over a simulation time of 1.3 μs and include glycerol molecules with at least one atom within 4 Å of Ile58 (green) of HEL. Red wireframes demark solvent regions with high local glycerol concentrations ($c_X(\vec{r})$ > $1.5c_{X,bulk}$).

Supporting References

1. Vagenende, V., M. G. S. Yap, and B. L. Trout. 2009. Mechanisms of Protein Stabilization and Prevention of Protein Aggregation by Glycerol. Biochemistry 48:11084-11096.