

Patterns, Volume 4

Supplemental information

**Optimal decision-making in high-throughput
virtual screening pipelines**

Hyun-Myung Woo, Xiaoning Qian, Li Tan, Shantenu Jha, Francis J. Alexander, Edward R. Dougherty, and Byung-Jun Yoon

Performance evaluations of the high-throughput virtual screening pipelines with various covariance matrices

Case 1: Uniform correlation between the surrogate models

$$p(y_1, y_2, y_3, y_4) \sim \mathcal{N} \left(\boldsymbol{\mu} = \mathbf{0}, \boldsymbol{\Sigma} = \begin{bmatrix} 1 & 0.2 & 0.2 & 0.2 \\ 0.2 & 1 & 0.2 & 0.2 \\ 0.2 & 0.2 & 1 & 0.2 \\ 0.2 & 0.2 & 0.2 & 1 \end{bmatrix} \right) \quad (1)$$

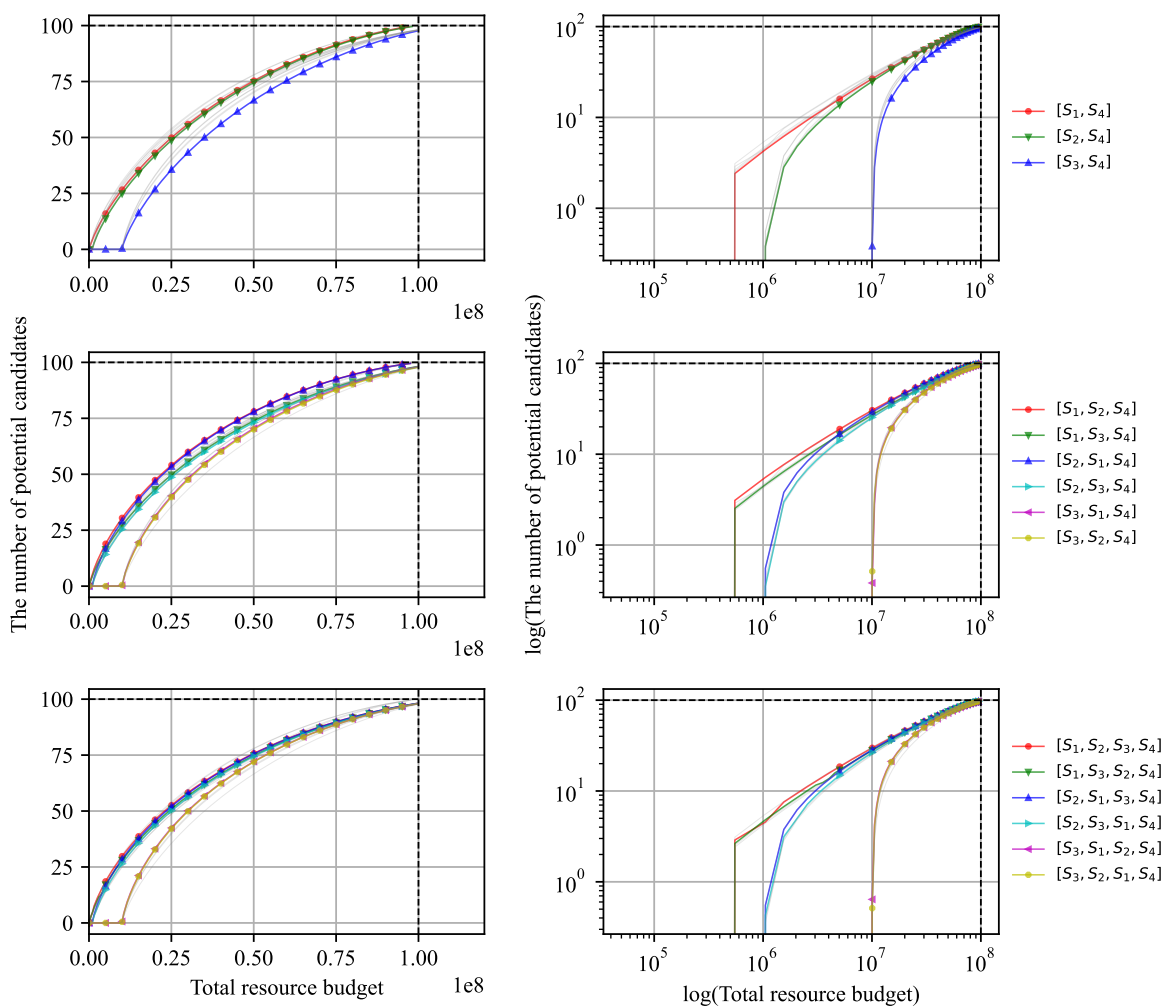


Figure S1: Performance comparison of the optimized high-throughput virtual screening pipelines in terms of discovery capability.

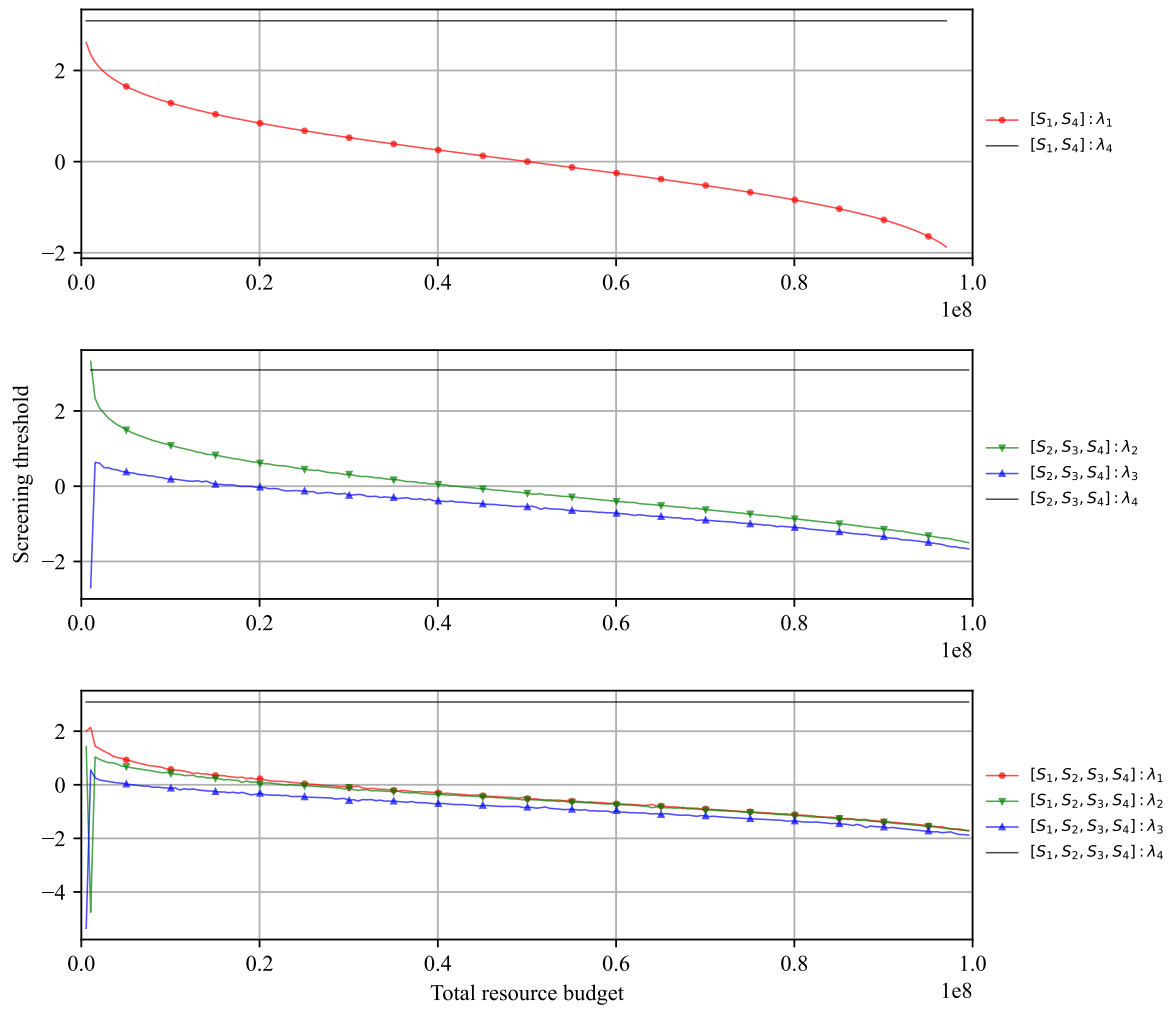


Figure S2: Screening thresholds of the optimized pipelines.

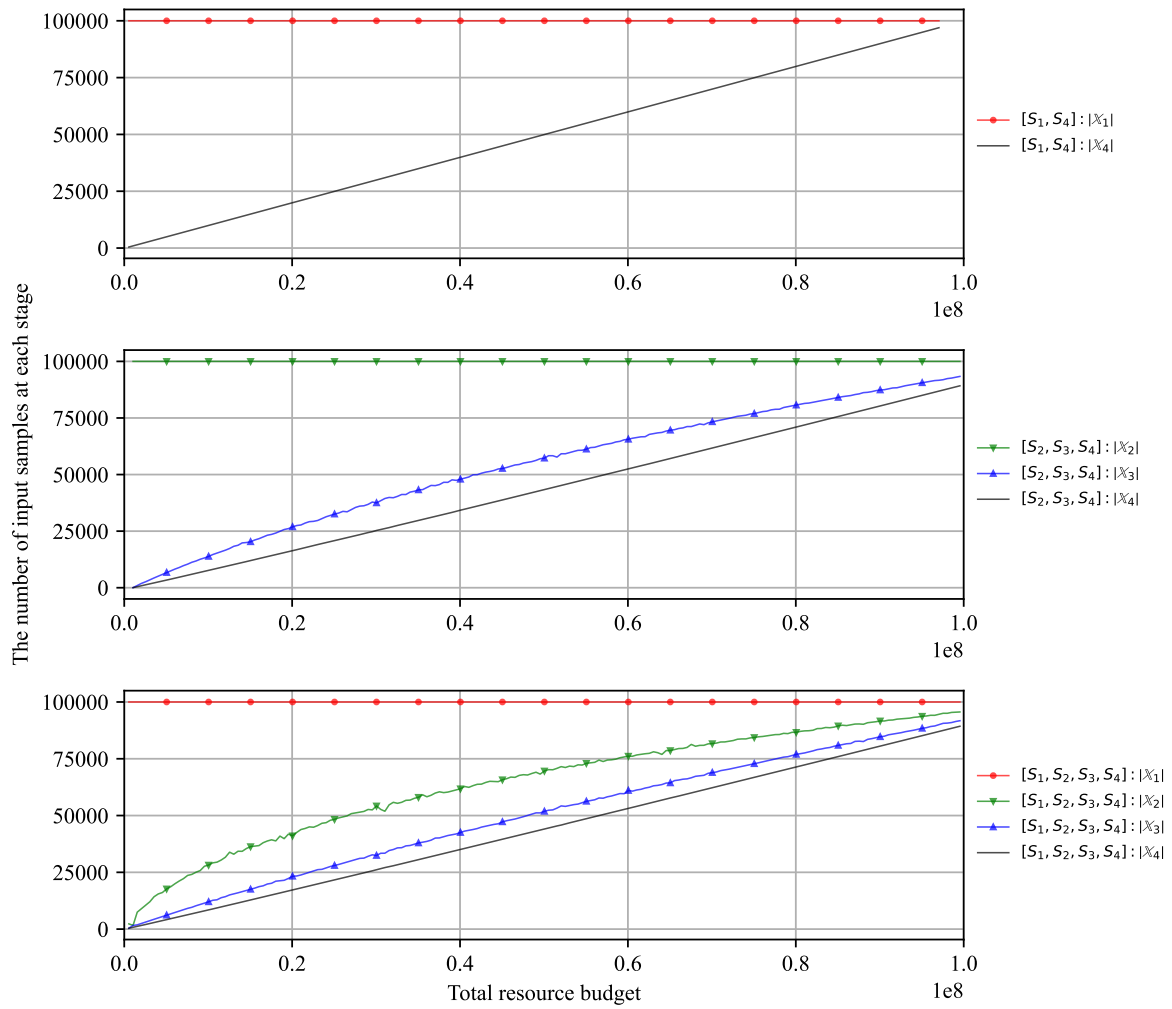


Figure S3: The number of input samples at each stage.

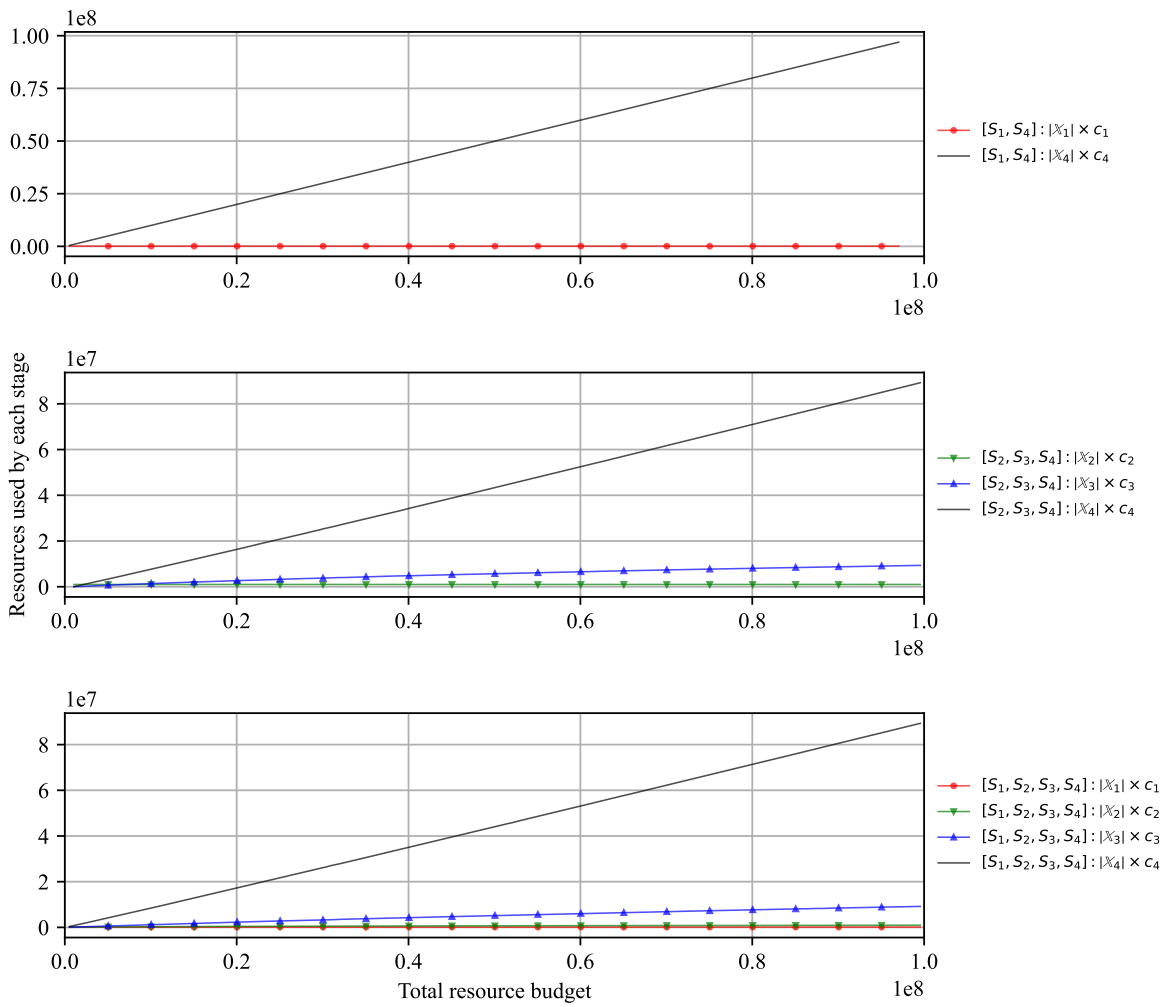


Figure S4: Resources used by each stage.

$$p(y_1, y_2, y_3, y_4) \sim \mathcal{N} \left(\boldsymbol{\mu} = \mathbf{0}, \boldsymbol{\Sigma} = \begin{bmatrix} 1 & 0.5 & 0.5 & 0.5 \\ 0.5 & 1 & 0.5 & 0.5 \\ 0.5 & 0.5 & 1 & 0.5 \\ 0.5 & 0.5 & 0.5 & 1 \end{bmatrix} \right) \quad (2)$$

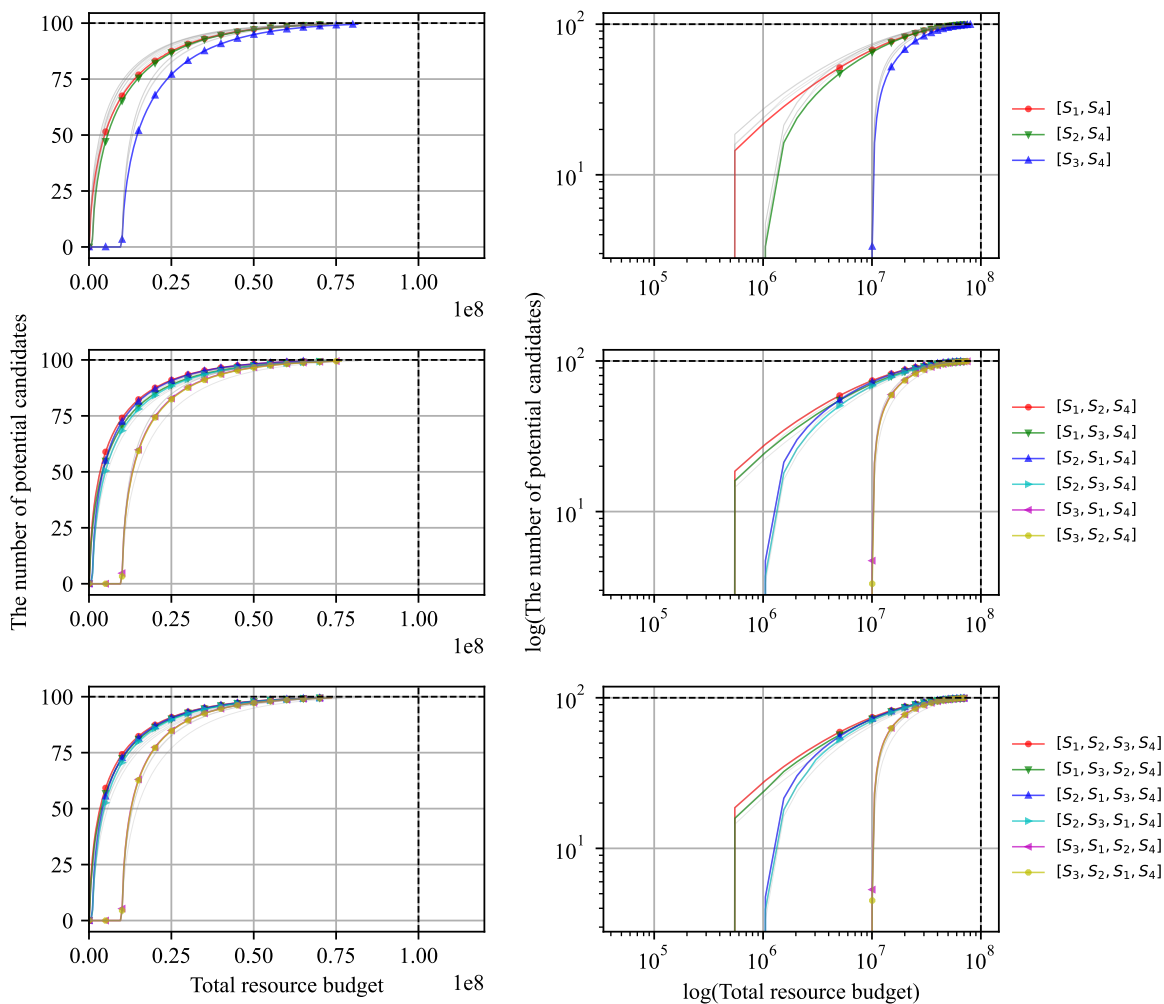


Figure S5: **Performance comparison of the optimized high-throughput virtual screening pipelines in terms of discovery capability.**

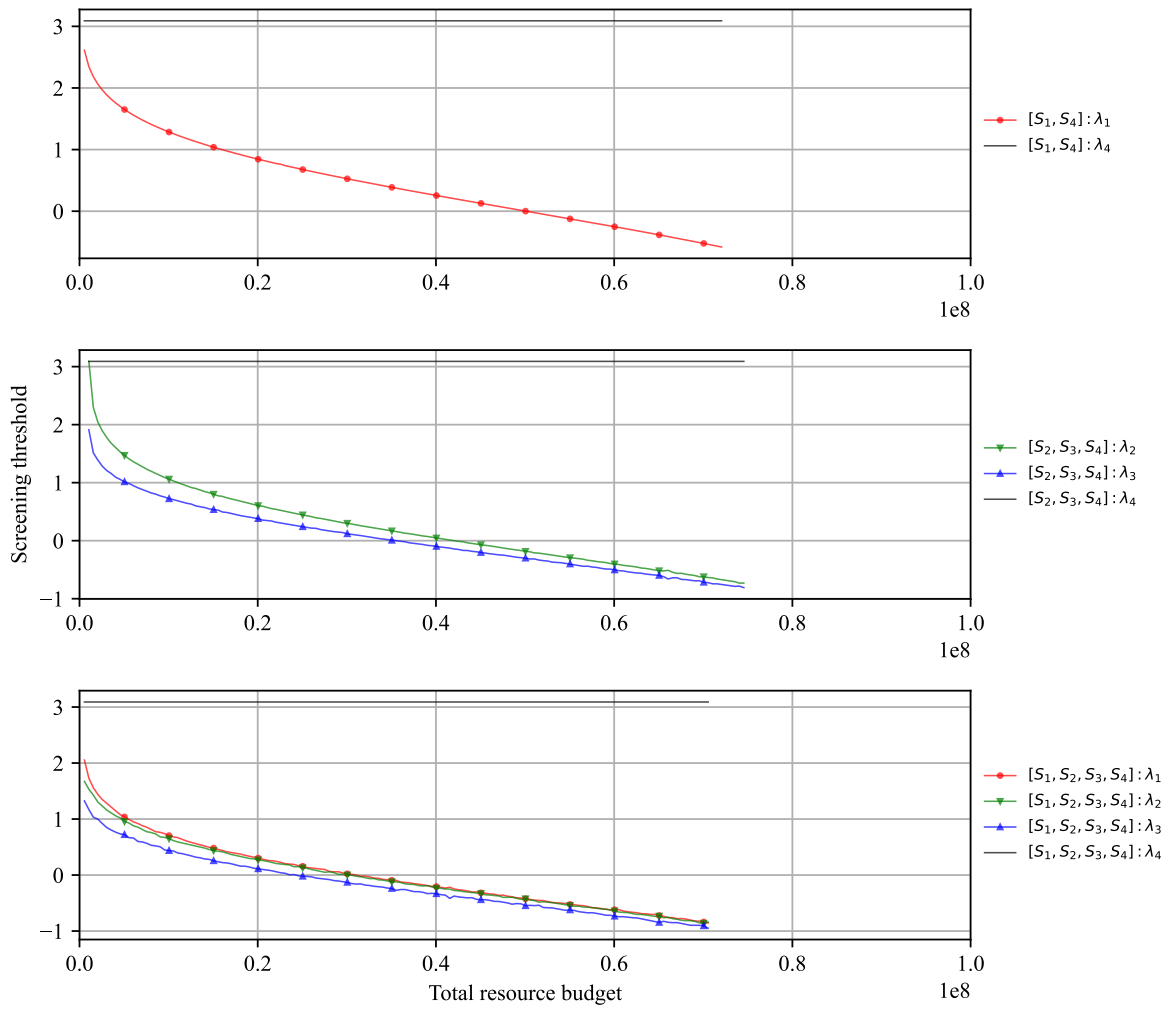


Figure S6: Screening thresholds of the optimized pipelines.

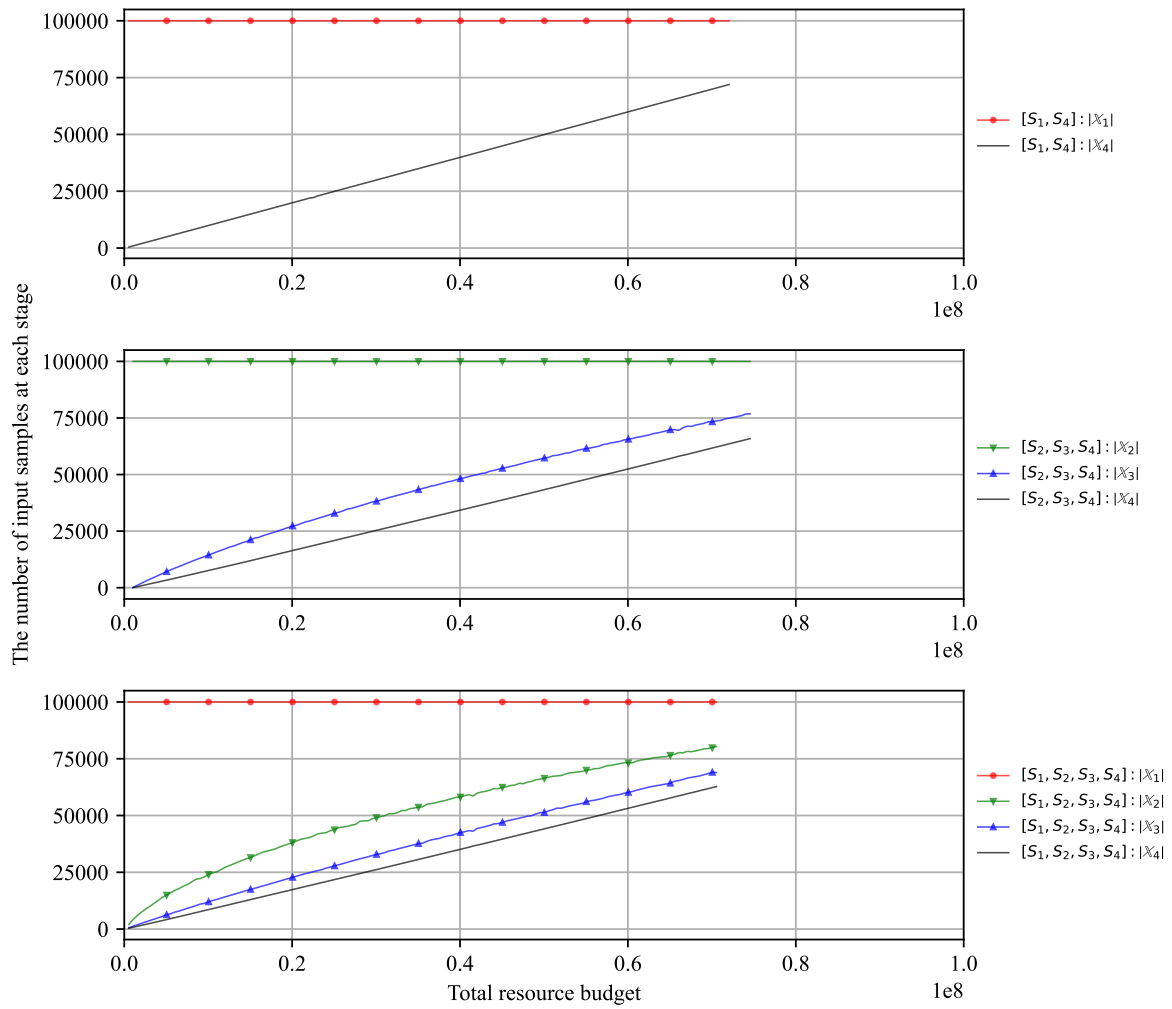


Figure S7: The number of input samples at each stage.

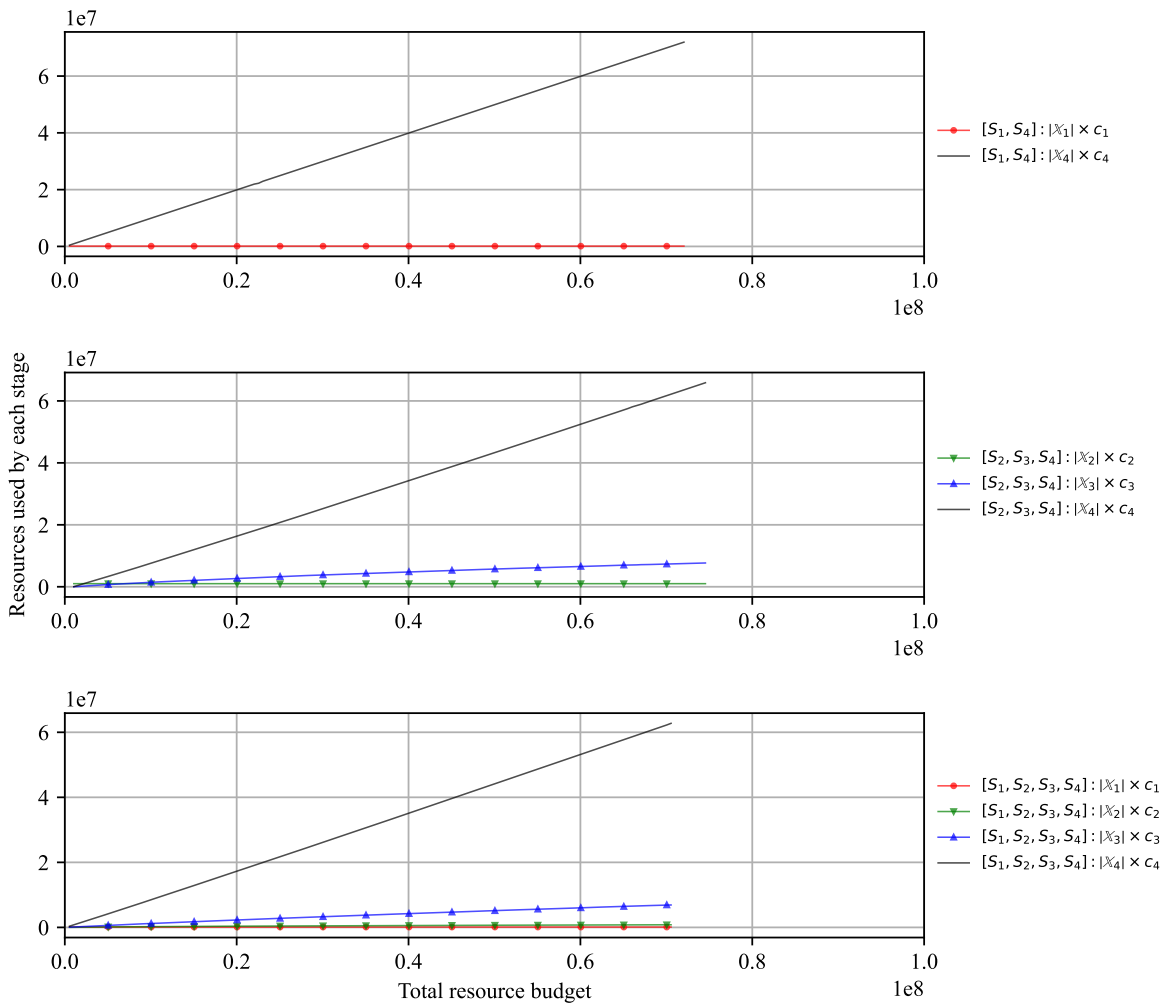


Figure S8: Resources used by each stage.

$$p(y_1, y_2, y_3, y_4) \sim \mathcal{N} \left(\boldsymbol{\mu} = \mathbf{0}, \boldsymbol{\Sigma} = \begin{bmatrix} 1 & 0.8 & 0.8 & 0.8 \\ 0.8 & 1 & 0.8 & 0.8 \\ 0.8 & 0.8 & 1 & 0.8 \\ 0.8 & 0.8 & 0.8 & 1 \end{bmatrix} \right) \quad (3)$$

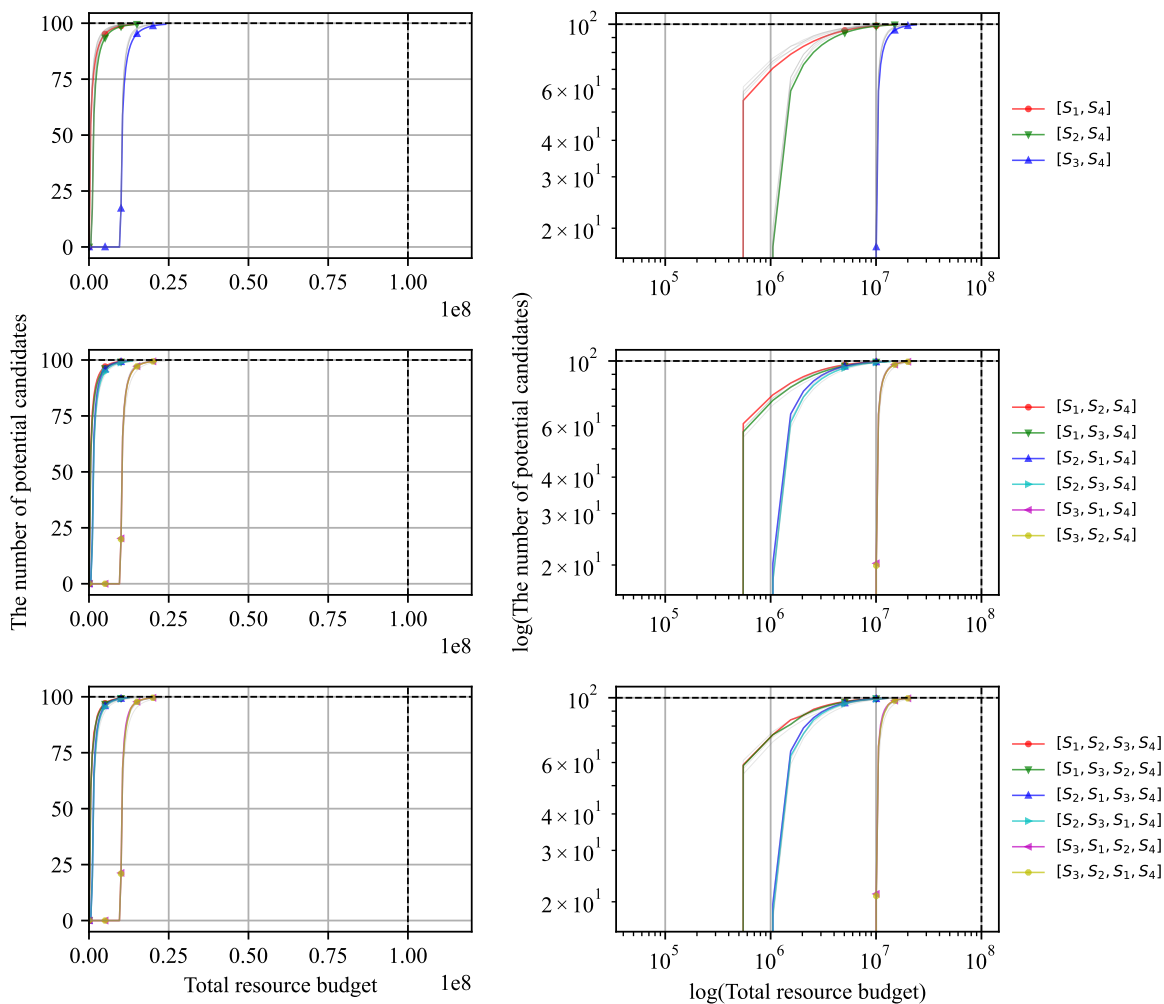


Figure S9: **Performance comparison of the optimized high-throughput virtual screening pipelines in terms of discovery capability.**

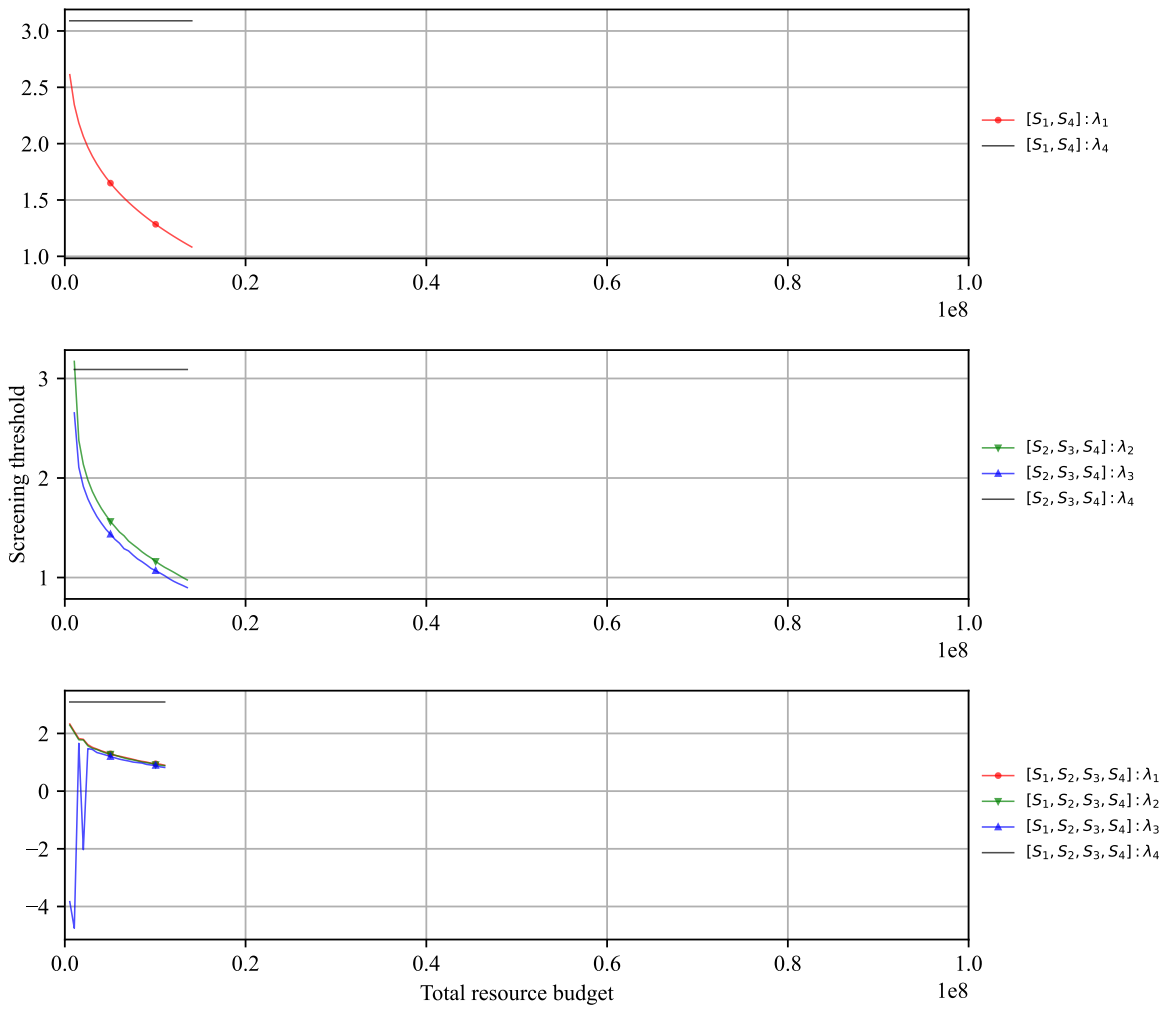


Figure S10: Screening thresholds of the optimized pipelines.

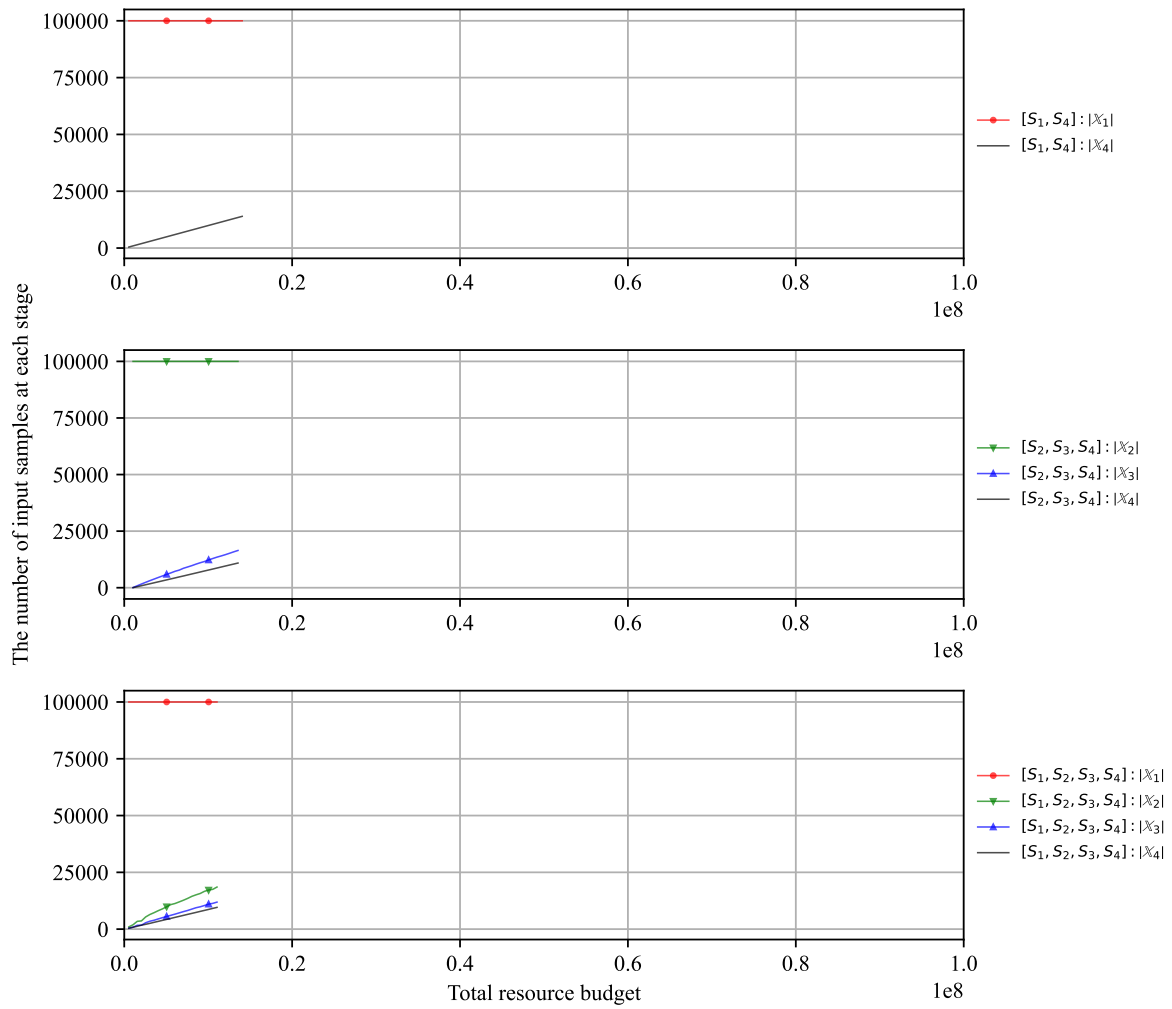


Figure S11: The number of input samples at each stage.

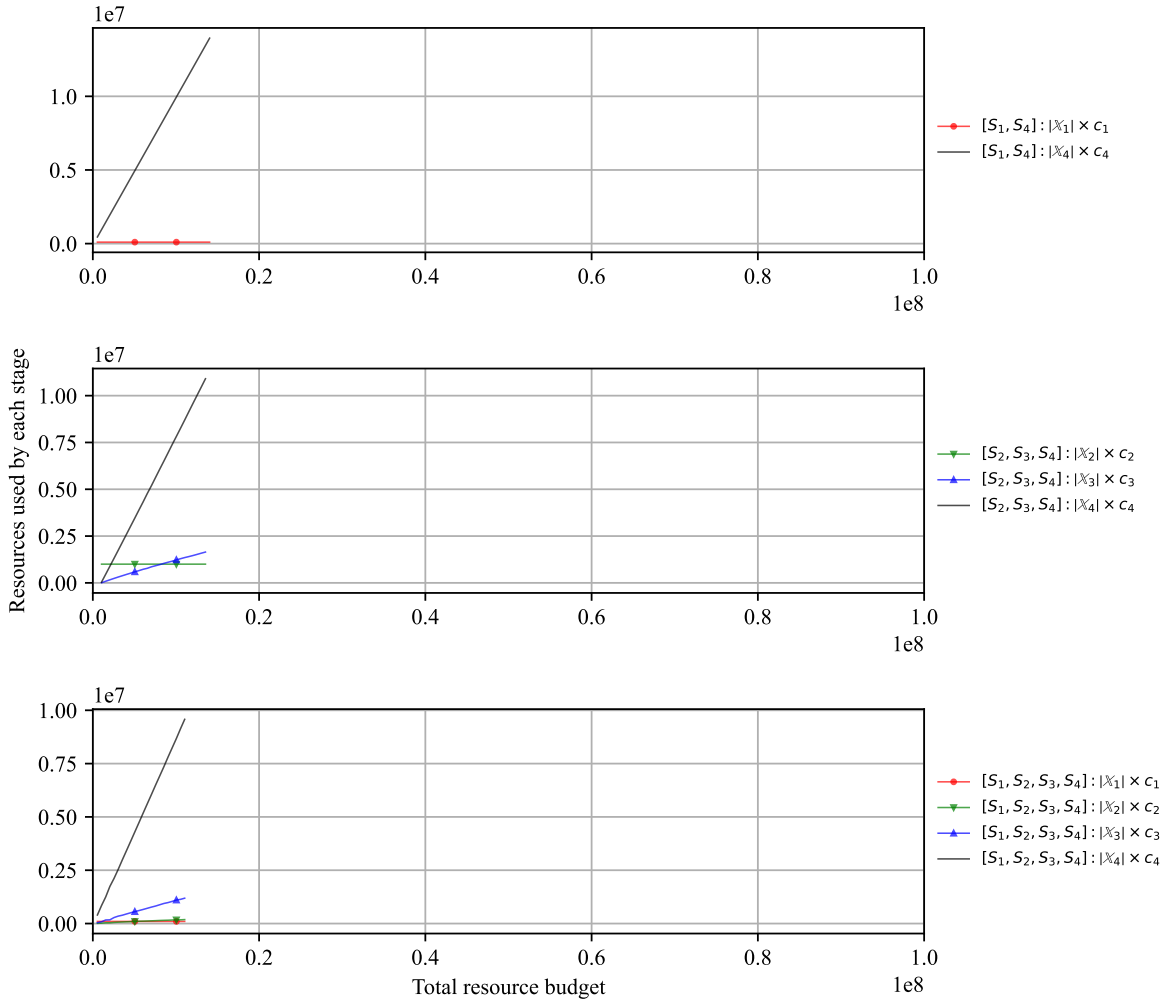


Figure S12: Resources used by each stage.

Case 2: Non-uniform correlation between the surrogate models

$$p(y_1, y_2, y_3, y_4) \sim \mathcal{N} \left(\boldsymbol{\mu} = \mathbf{0}, \boldsymbol{\Sigma} = \begin{bmatrix} 1 & 0.2 & 0.2 & 0.8 \\ 0.2 & 1 & 0.2 & 0.2 \\ 0.2 & 0.2 & 1 & 0.2 \\ 0.8 & 0.2 & 0.2 & 1 \end{bmatrix} \right) \quad (4)$$

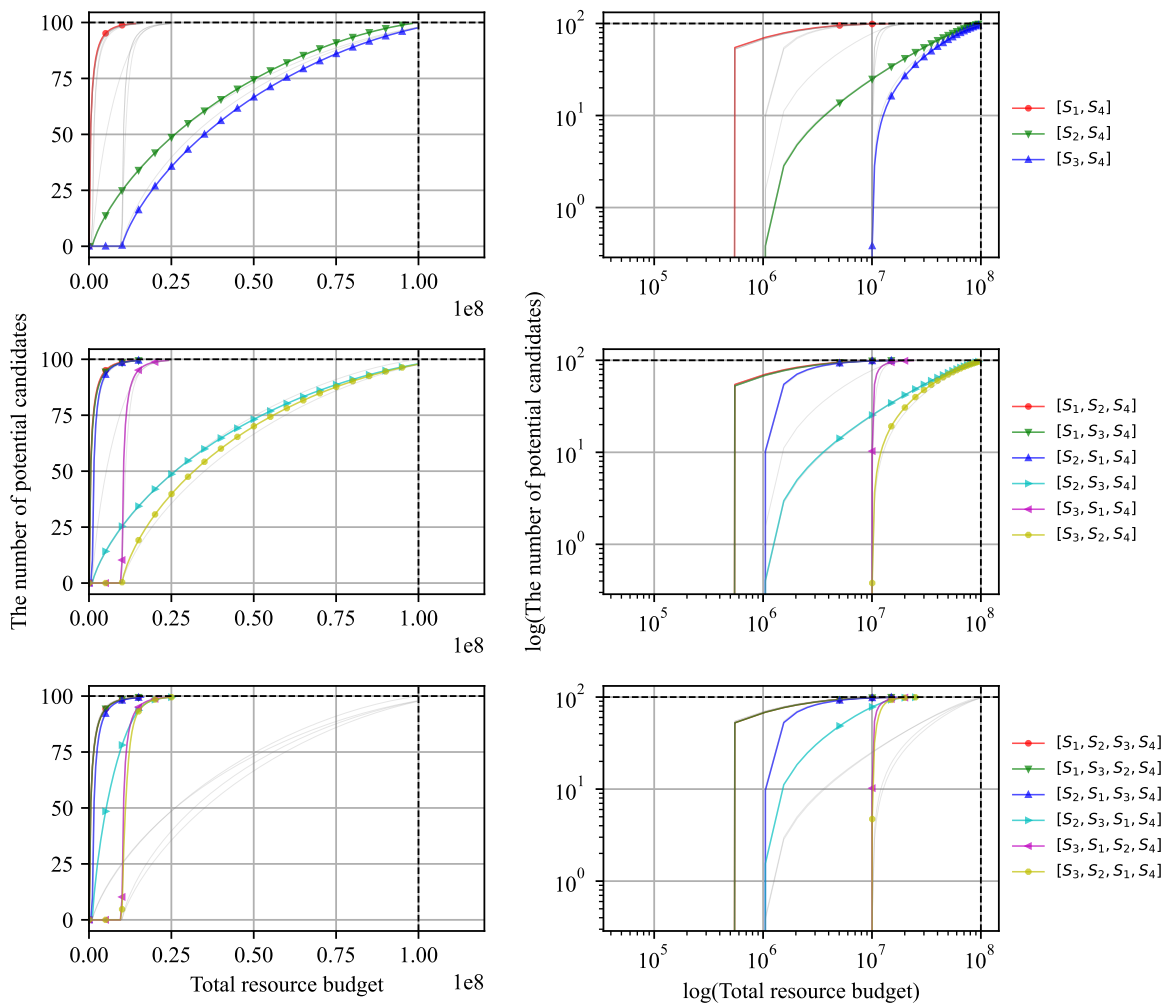


Figure S13: Performance comparison of the optimized high-throughput virtual screening pipelines in terms of discovery capability.

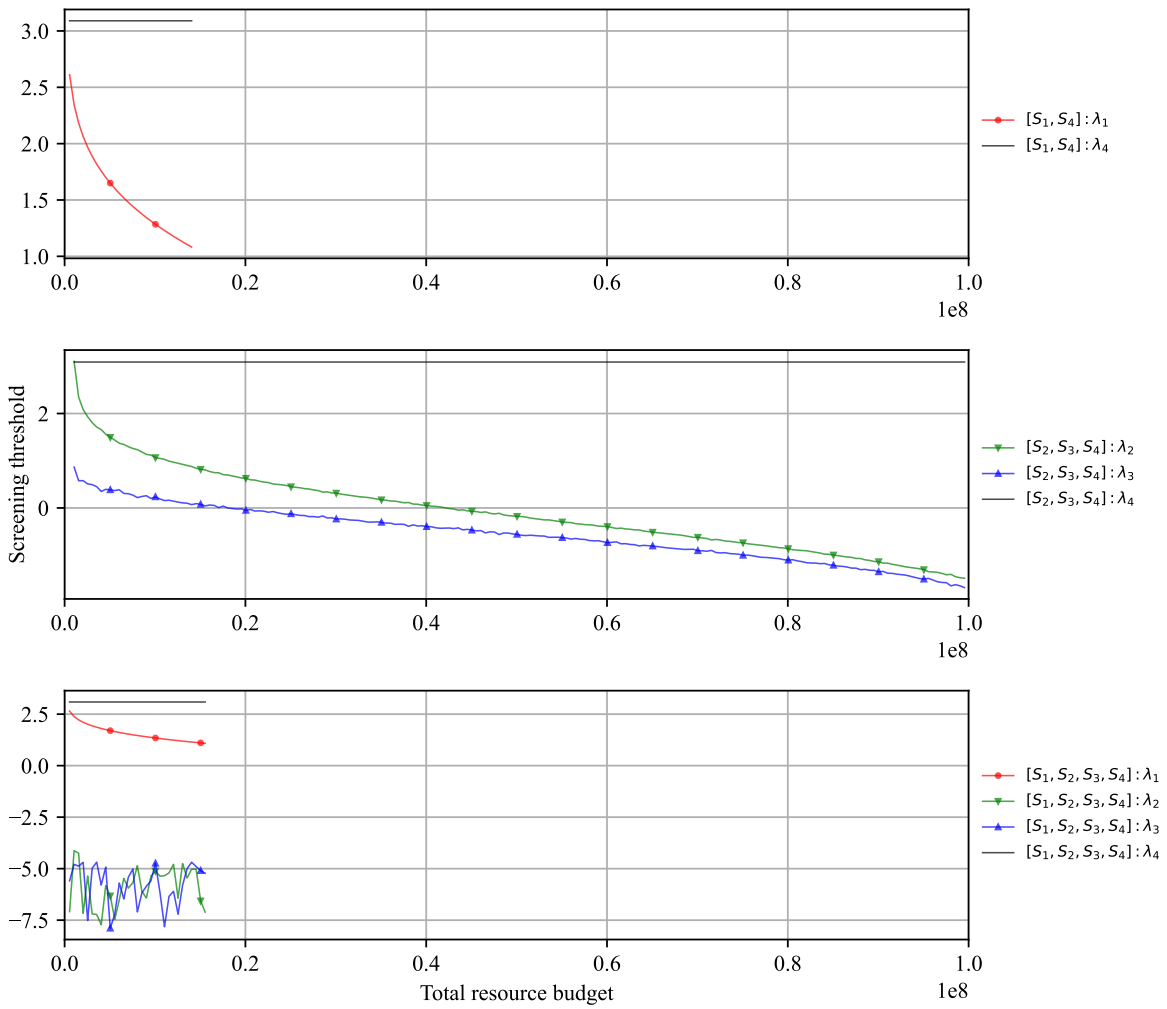


Figure S14: Screening thresholds of the optimized pipelines.

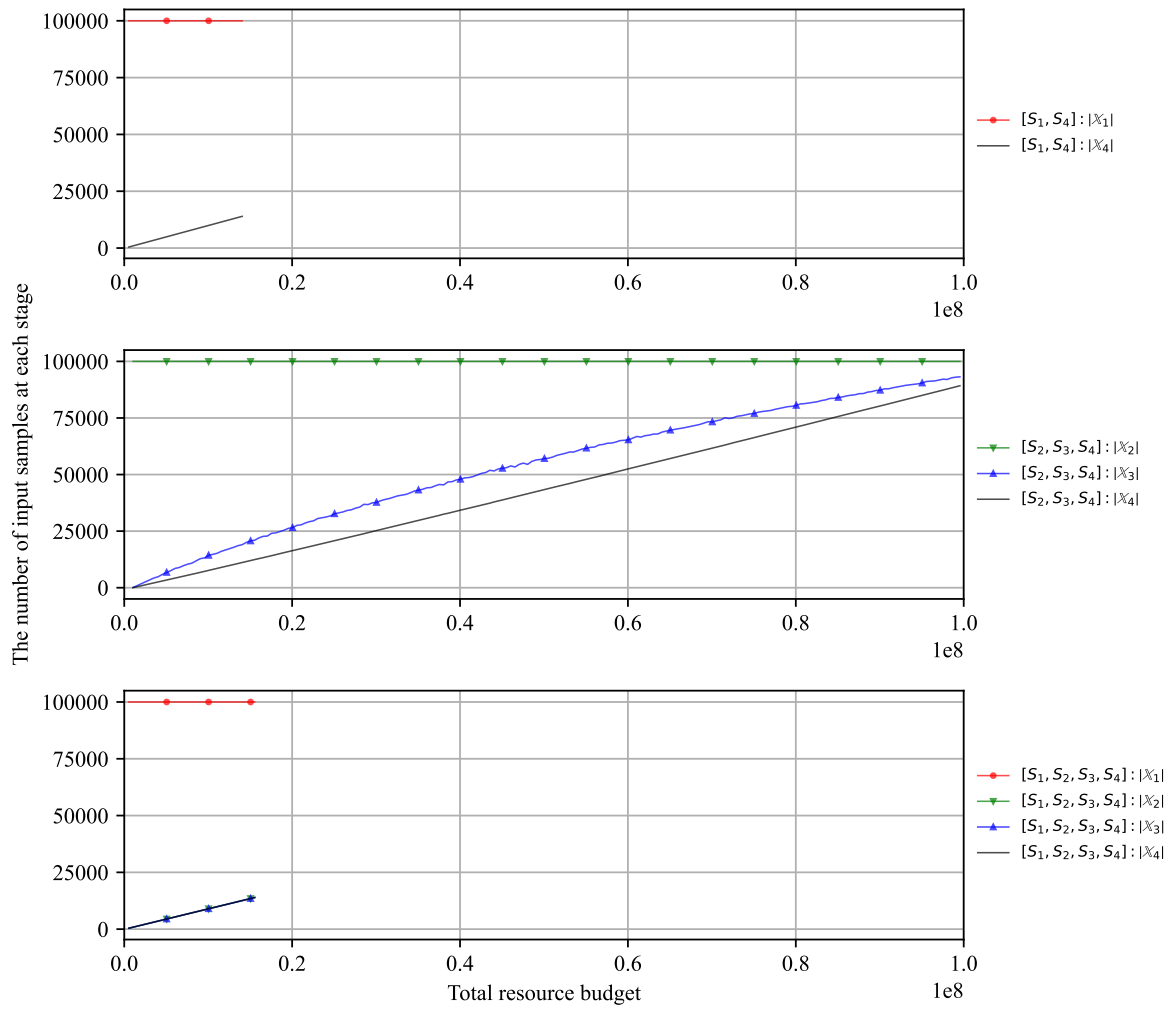


Figure S15: The number of input samples at each stage.

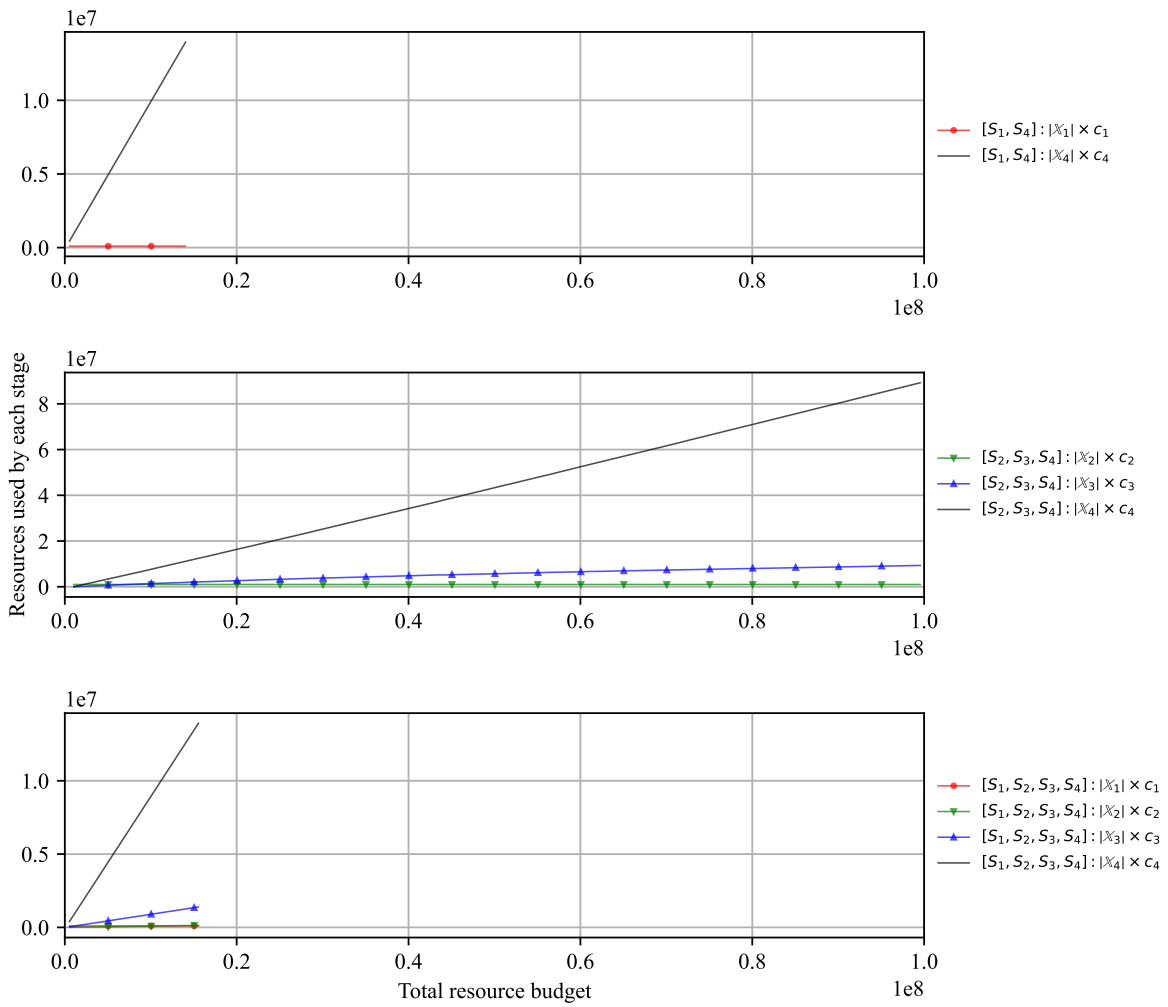


Figure S16: Resources used by each stage.

$$p(y_1, y_2, y_3, y_4) \sim \mathcal{N} \left(\boldsymbol{\mu} = \mathbf{0}, \boldsymbol{\Sigma} = \begin{bmatrix} 1 & 0.2 & 0.2 & 0.2 \\ 0.2 & 1 & 0.2 & 0.8 \\ 0.2 & 0.2 & 1 & 0.2 \\ 0.2 & 0.8 & 0.2 & 1 \end{bmatrix} \right) \quad (5)$$

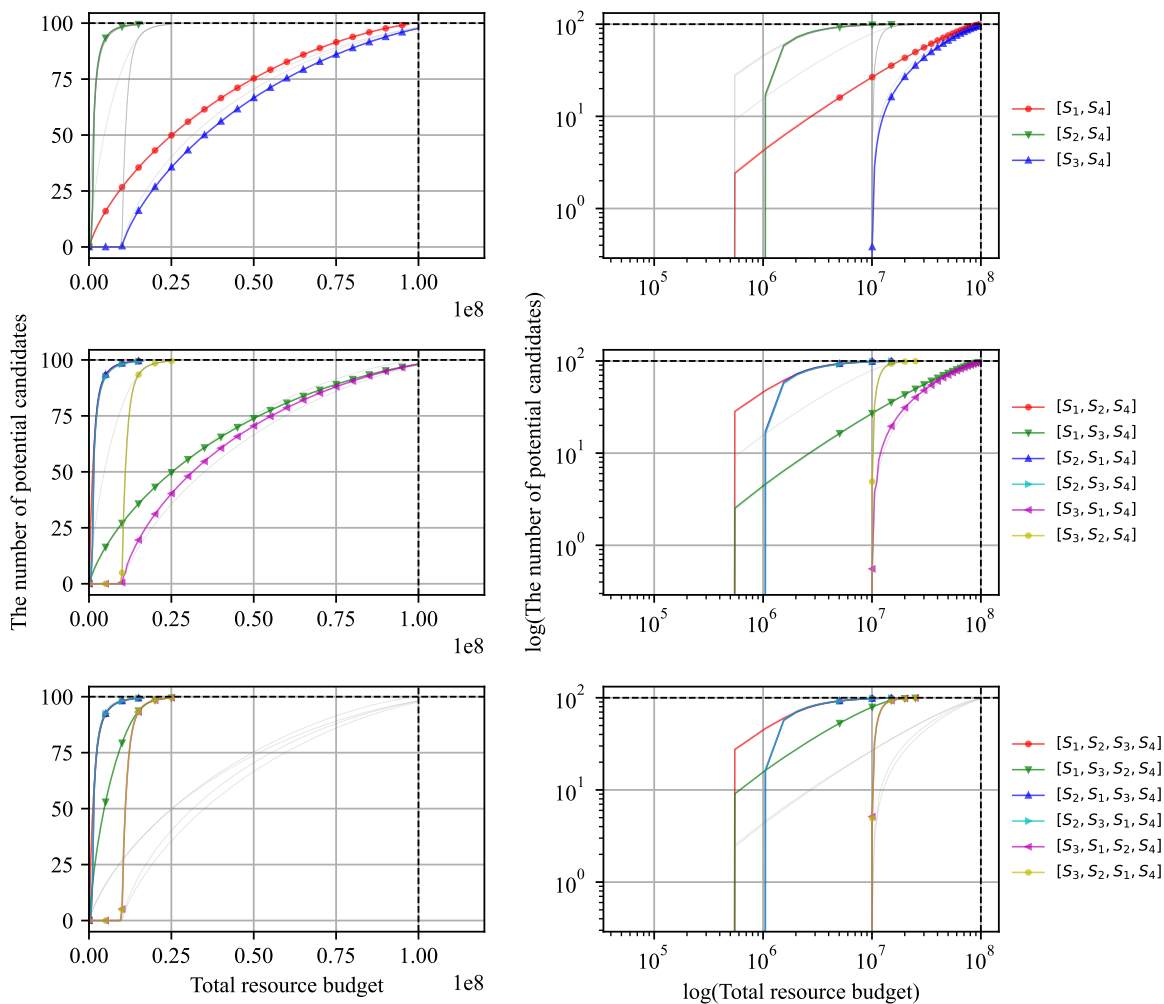


Figure S17: Performance comparison of the optimized high-throughput virtual screening pipelines in terms of discovery capability.

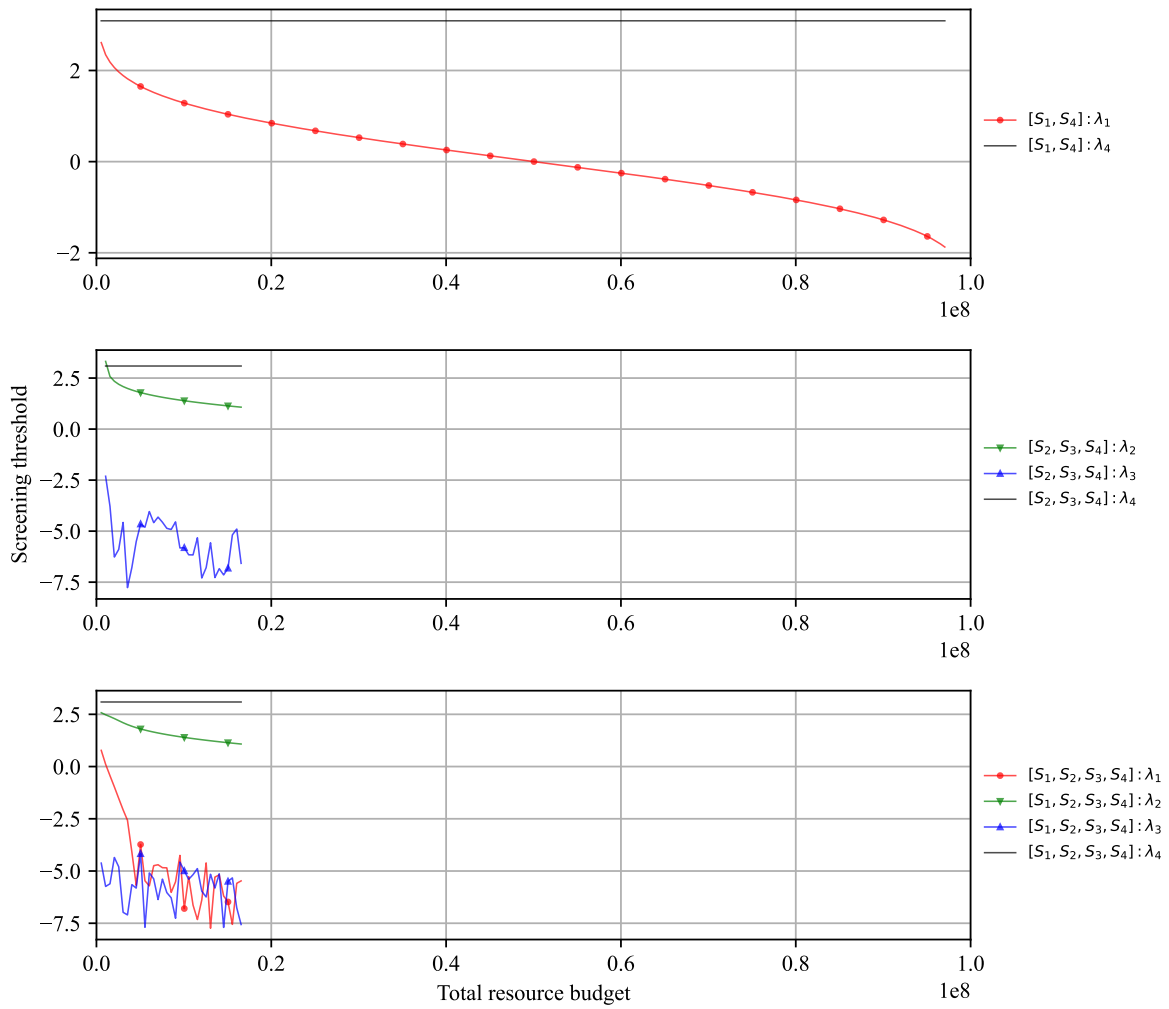


Figure S18: Screening thresholds of the optimized pipelines.

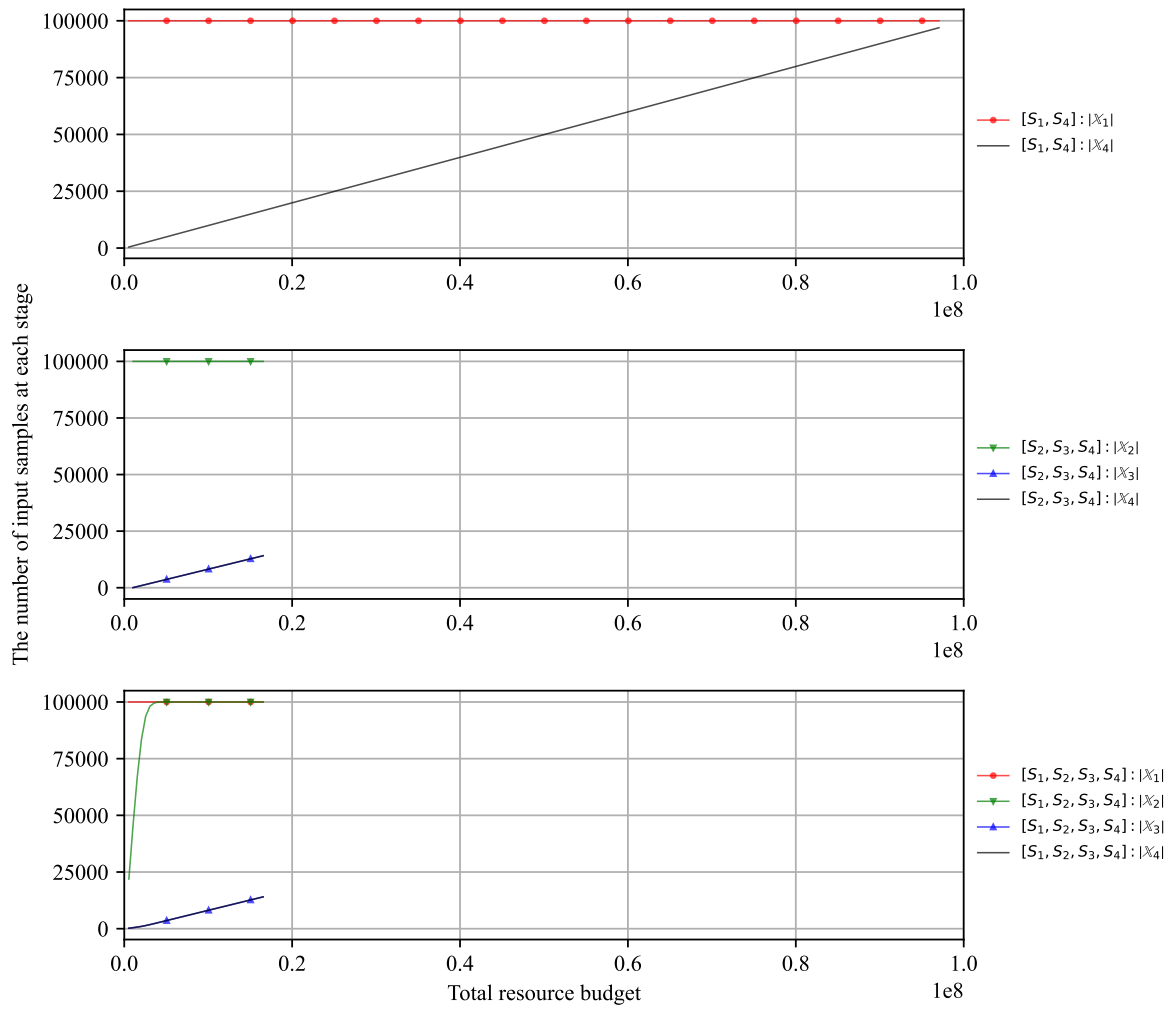


Figure S19: The number of input samples at each stage.

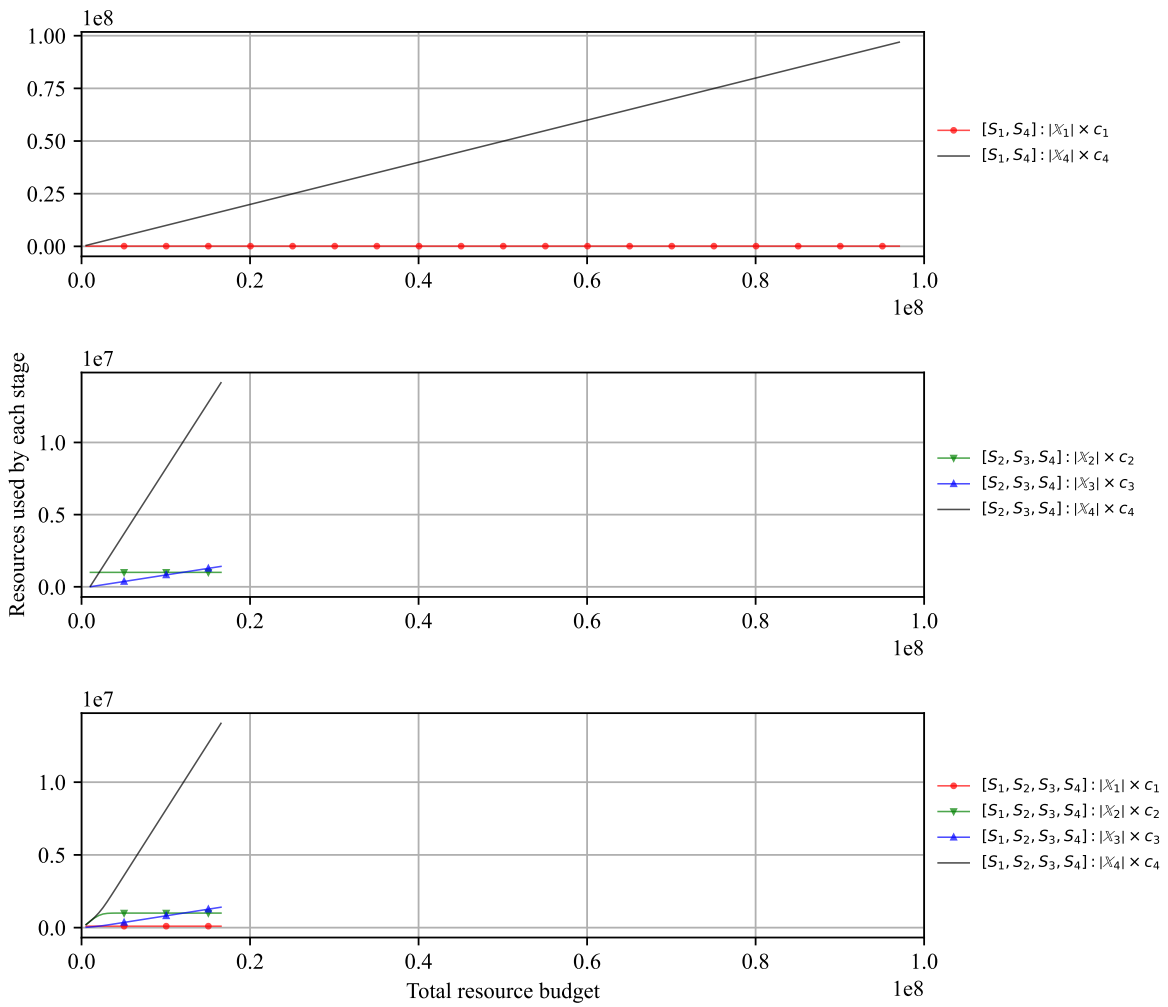


Figure S20: Resources used by each stage.

$$p(y_1, y_2, y_3, y_4) \sim \mathcal{N} \left(\boldsymbol{\mu} = \mathbf{0}, \boldsymbol{\Sigma} = \begin{bmatrix} 1 & 0.2 & 0.2 & 0.2 \\ 0.2 & 1 & 0.2 & 0.2 \\ 0.2 & 0.2 & 1 & 0.8 \\ 0.2 & 0.2 & 0.8 & 1 \end{bmatrix} \right) \quad (6)$$

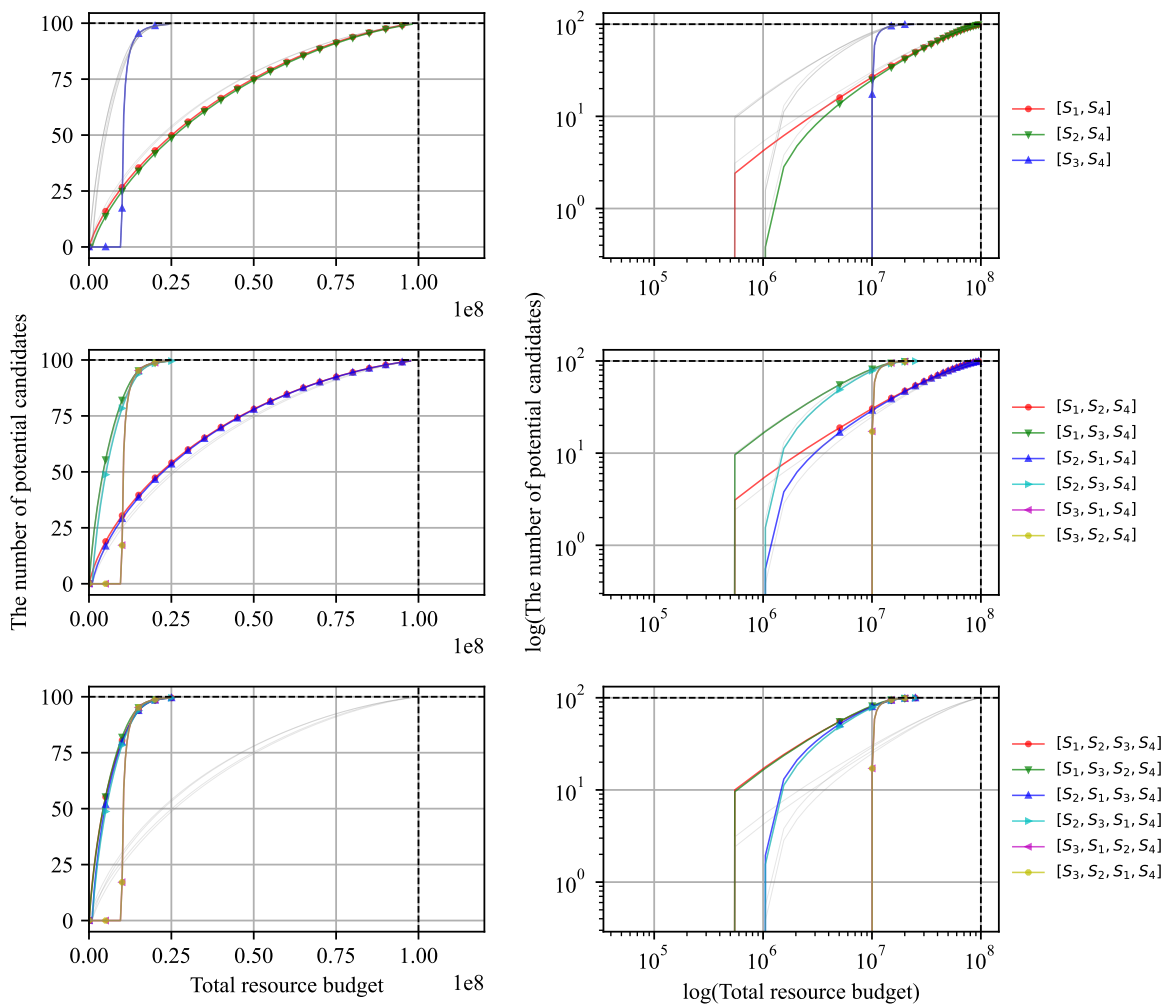


Figure S21: Performance comparison of the optimized high-throughput virtual screening pipelines in terms of discovery capability.

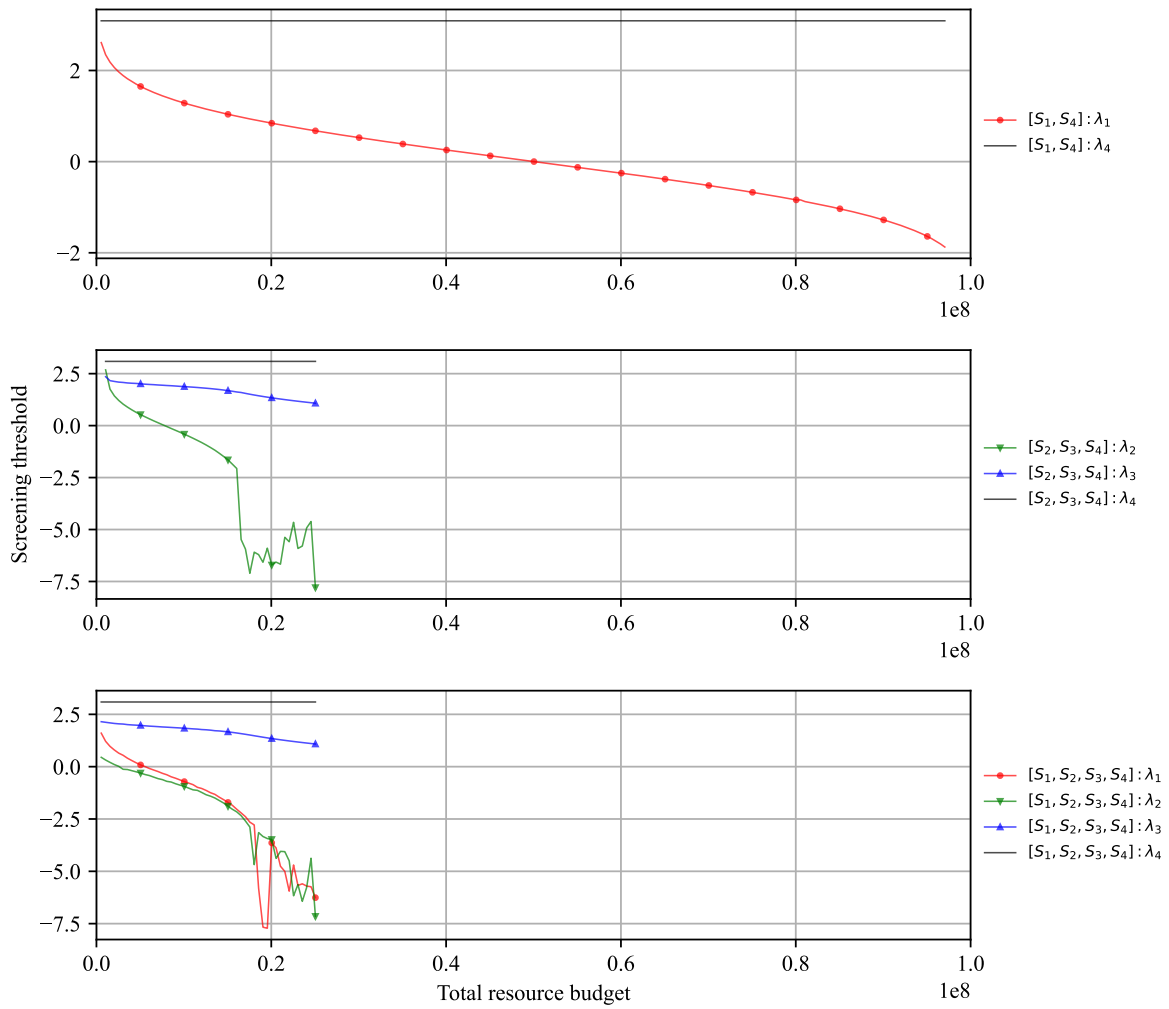


Figure S22: Screening thresholds of the optimized pipelines.

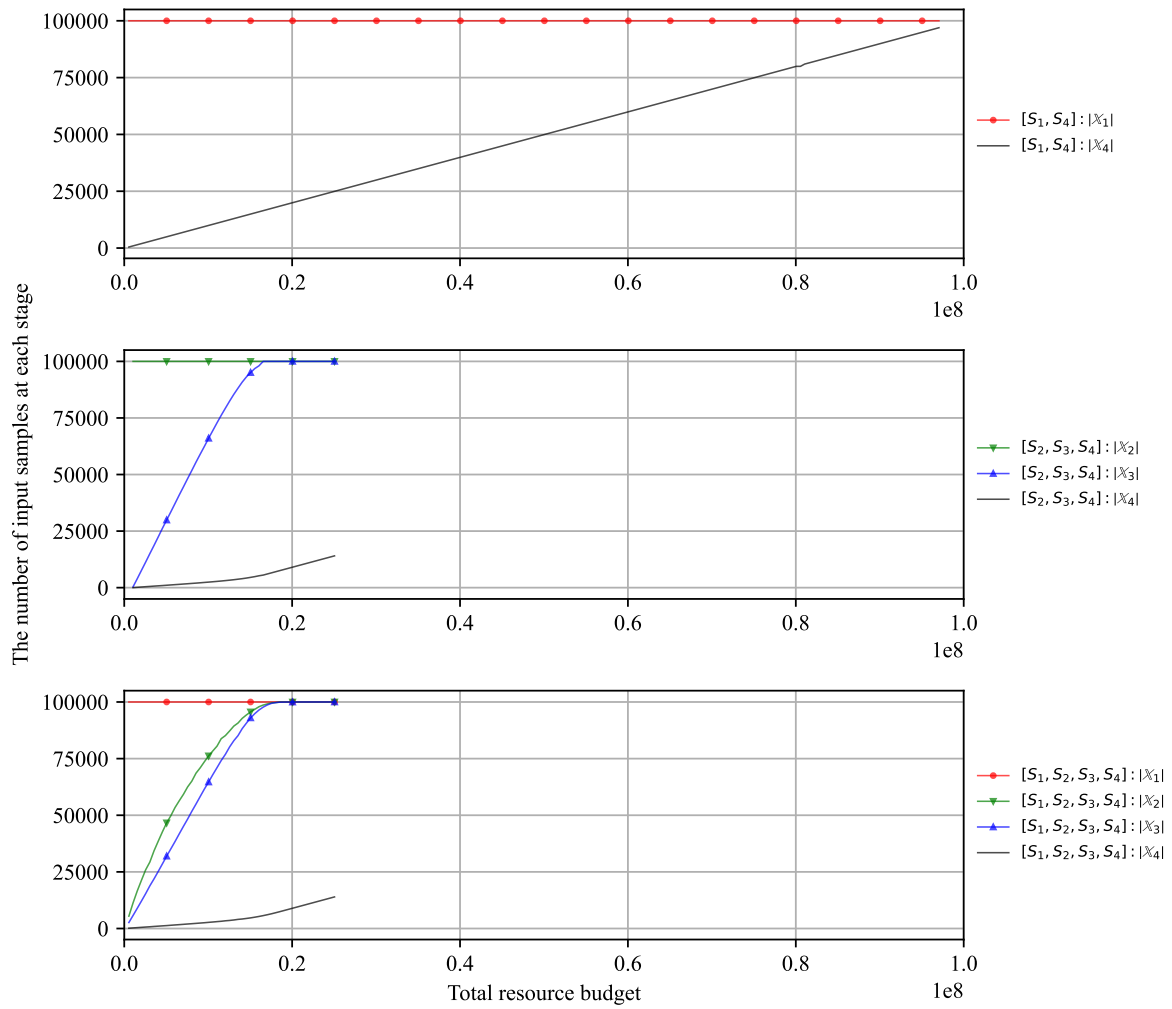


Figure S23: The number of input samples at each stage.

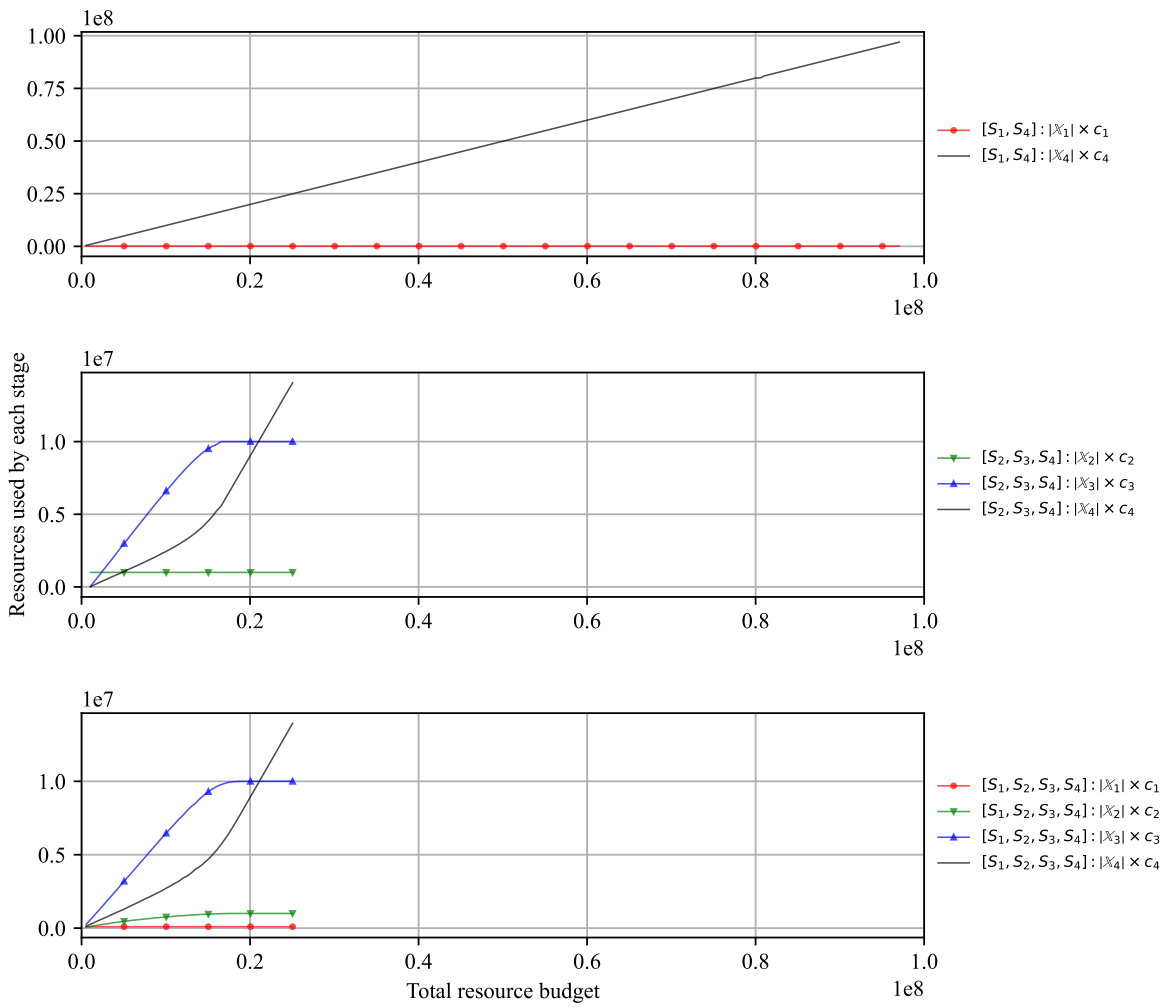


Figure S24: Resources used by each stage.

$$p(y_1, y_2, y_3, y_4) \sim \mathcal{N} \left(\boldsymbol{\mu} = \mathbf{0}, \boldsymbol{\Sigma} = \begin{bmatrix} 1 & 0.2 & 0.2 & 0.2 \\ 0.2 & 1 & 0.8 & 0.8 \\ 0.2 & 0.8 & 1 & 0.8 \\ 0.2 & 0.8 & 0.8 & 1 \end{bmatrix} \right) \quad (7)$$

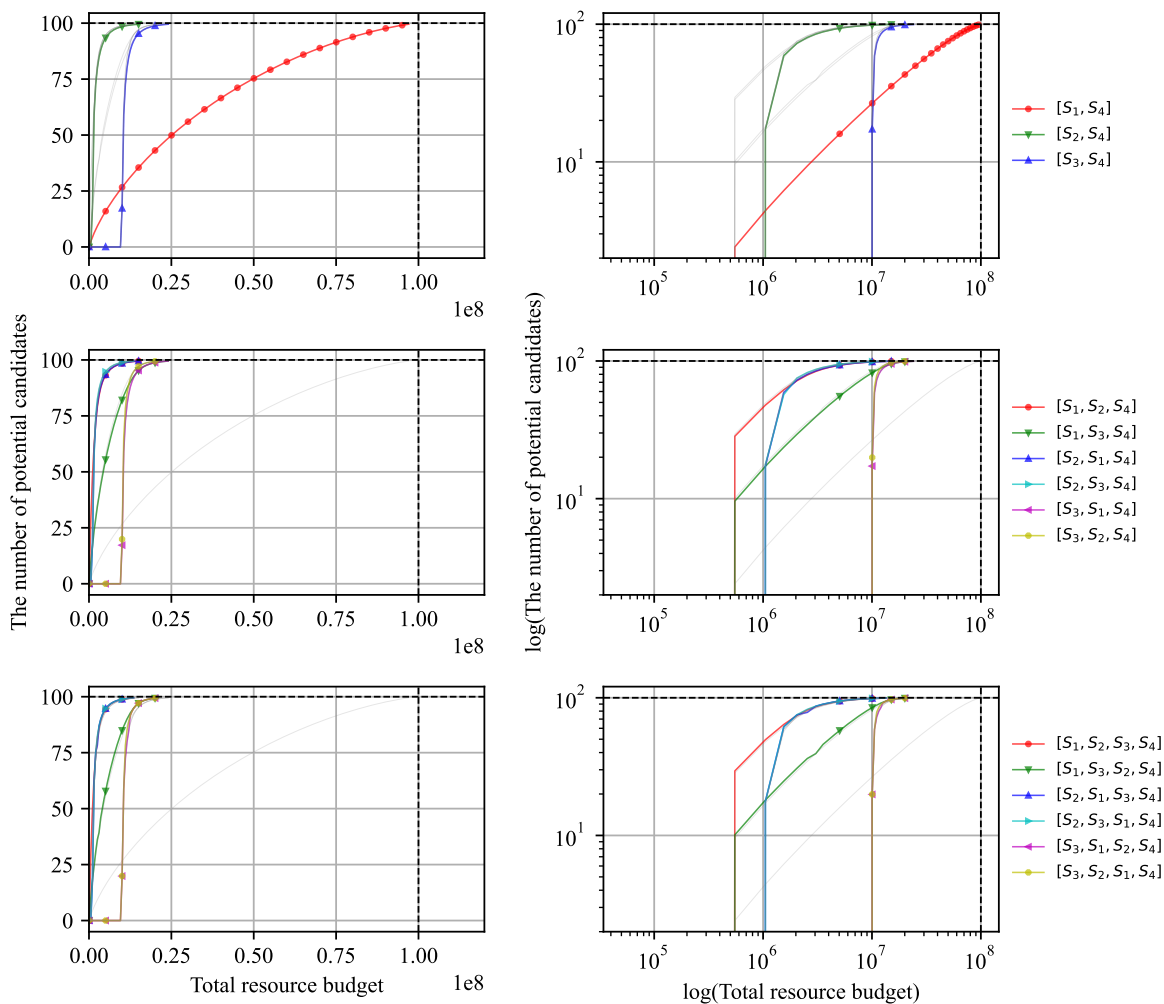


Figure S25: Performance comparison of the optimized high-throughput virtual screening pipelines in terms of discovery capability.

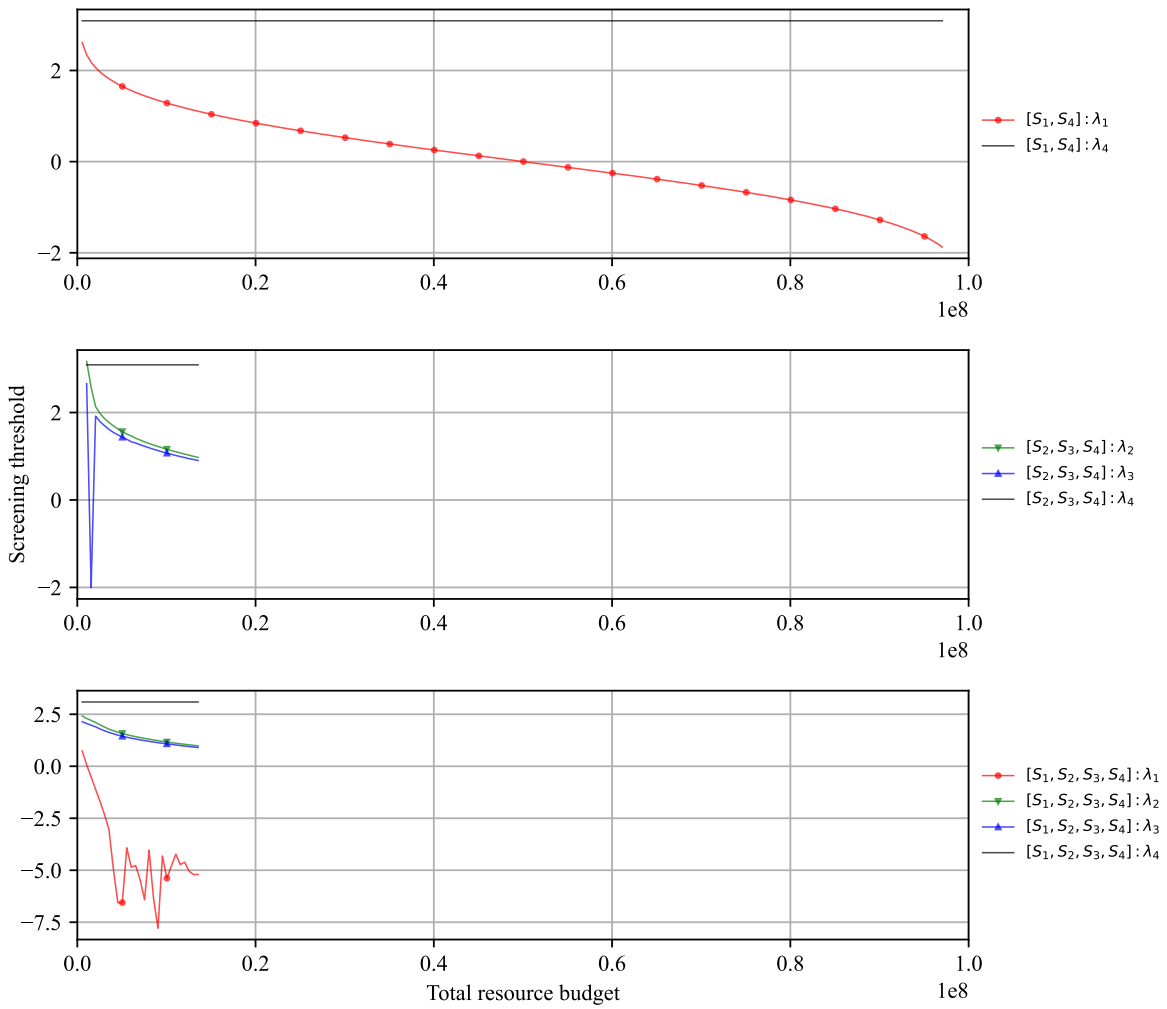


Figure S26: Screening thresholds of the optimized pipelines.

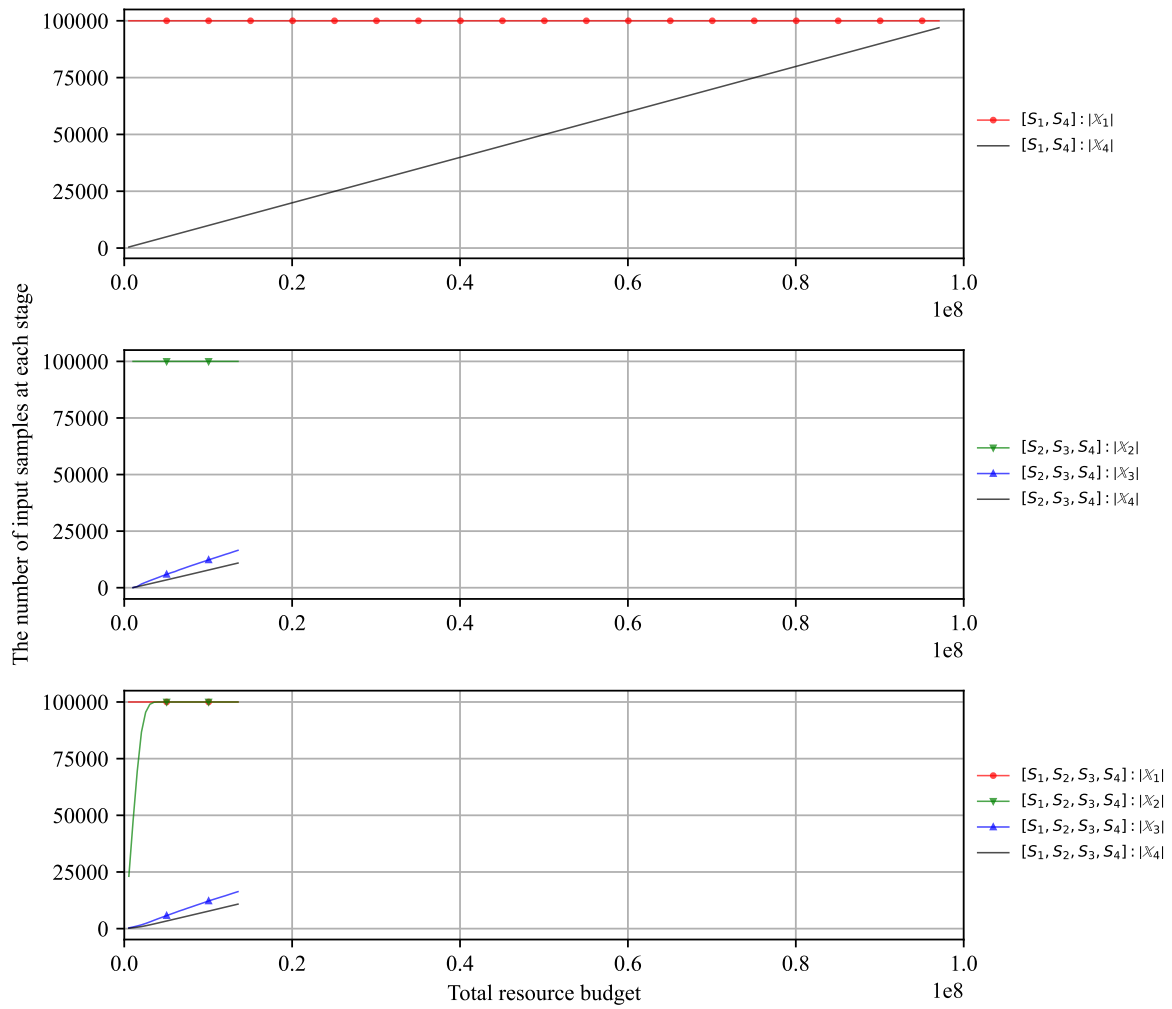


Figure S27: The number of input samples at each stage.

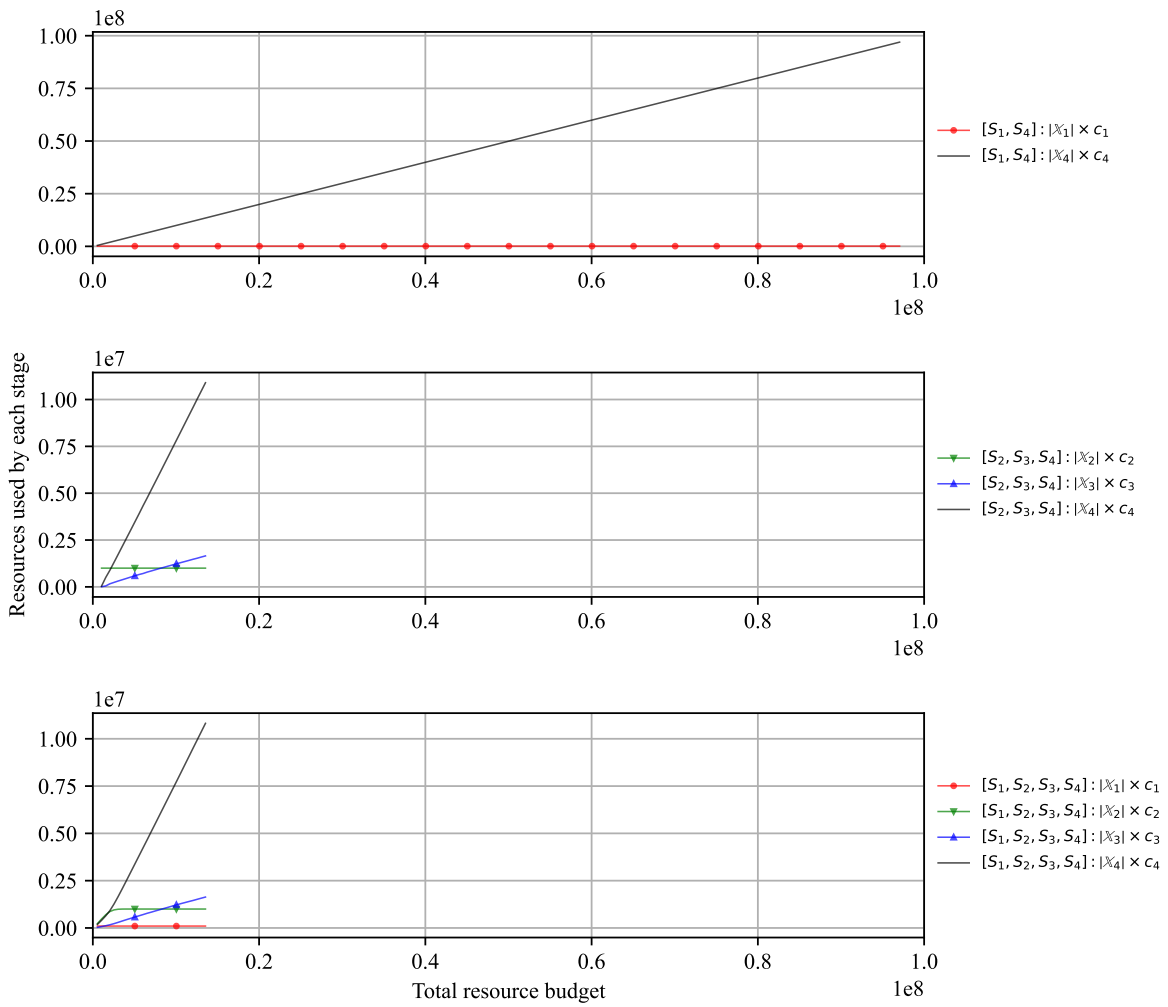


Figure S28: Resources used by each stage.

$$p(y_1, y_2, y_3, y_4) \sim \mathcal{N} \left(\boldsymbol{\mu} = \mathbf{0}, \boldsymbol{\Sigma} = \begin{bmatrix} 1 & 0.2 & 0.8 & 0.8 \\ 0.2 & 1 & 0.2 & 0.2 \\ 0.8 & 0.2 & 1 & 0.8 \\ 0.8 & 0.2 & 0.8 & 1 \end{bmatrix} \right) \quad (8)$$

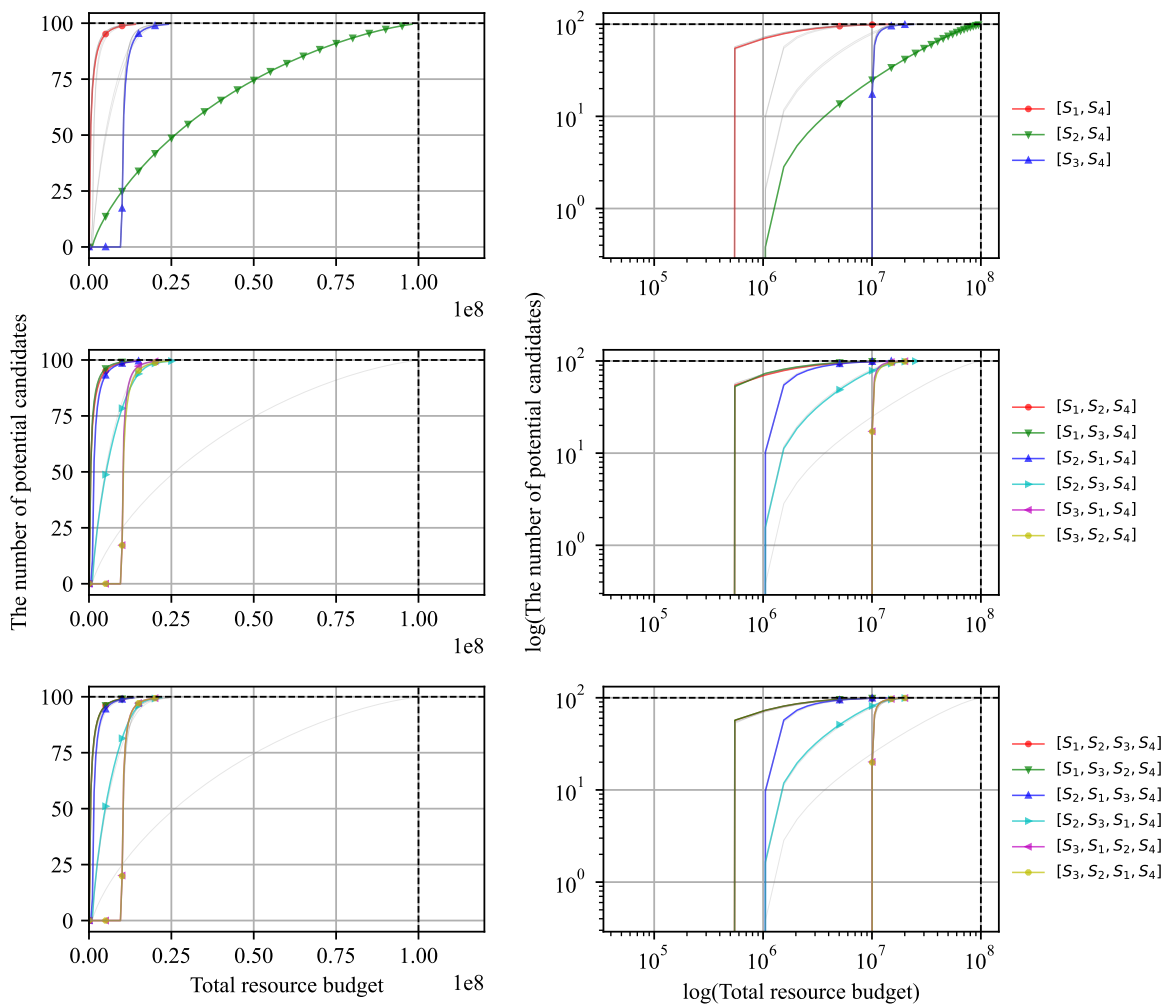


Figure S29: Performance comparison of the optimized high-throughput virtual screening pipelines in terms of discovery capability.

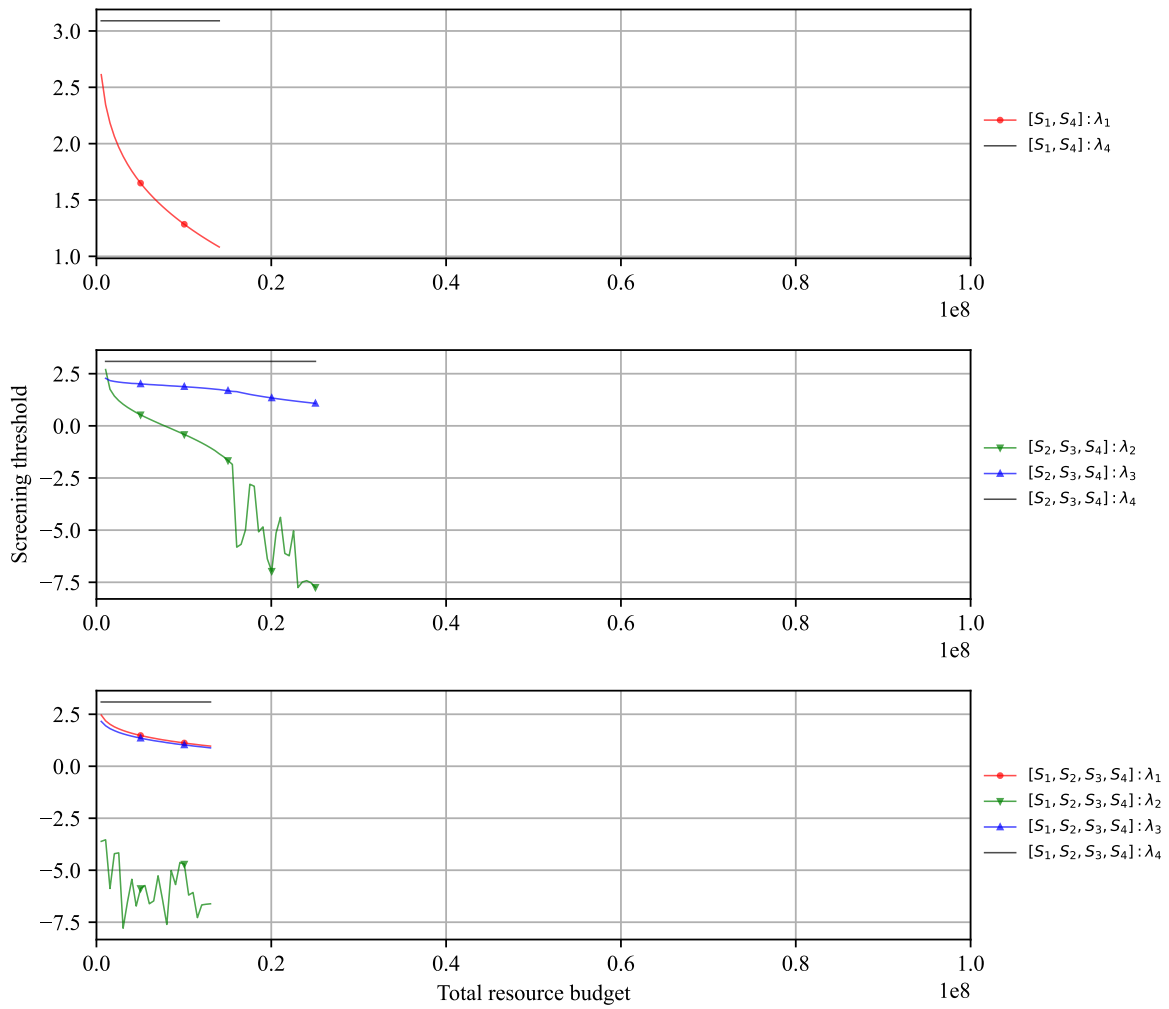


Figure S30: Screening thresholds of the optimized pipelines.

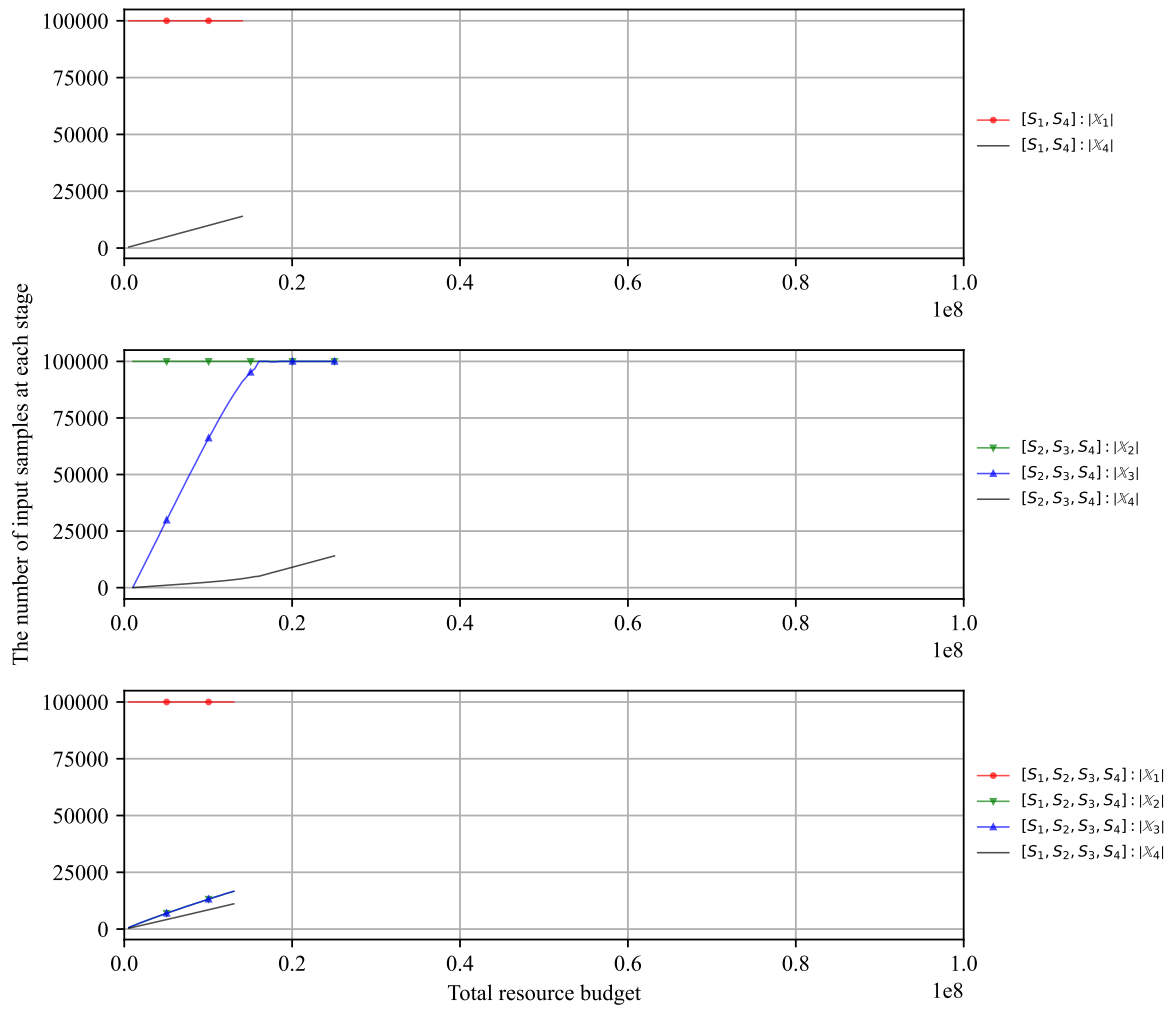


Figure S31: The number of input samples at each stage.

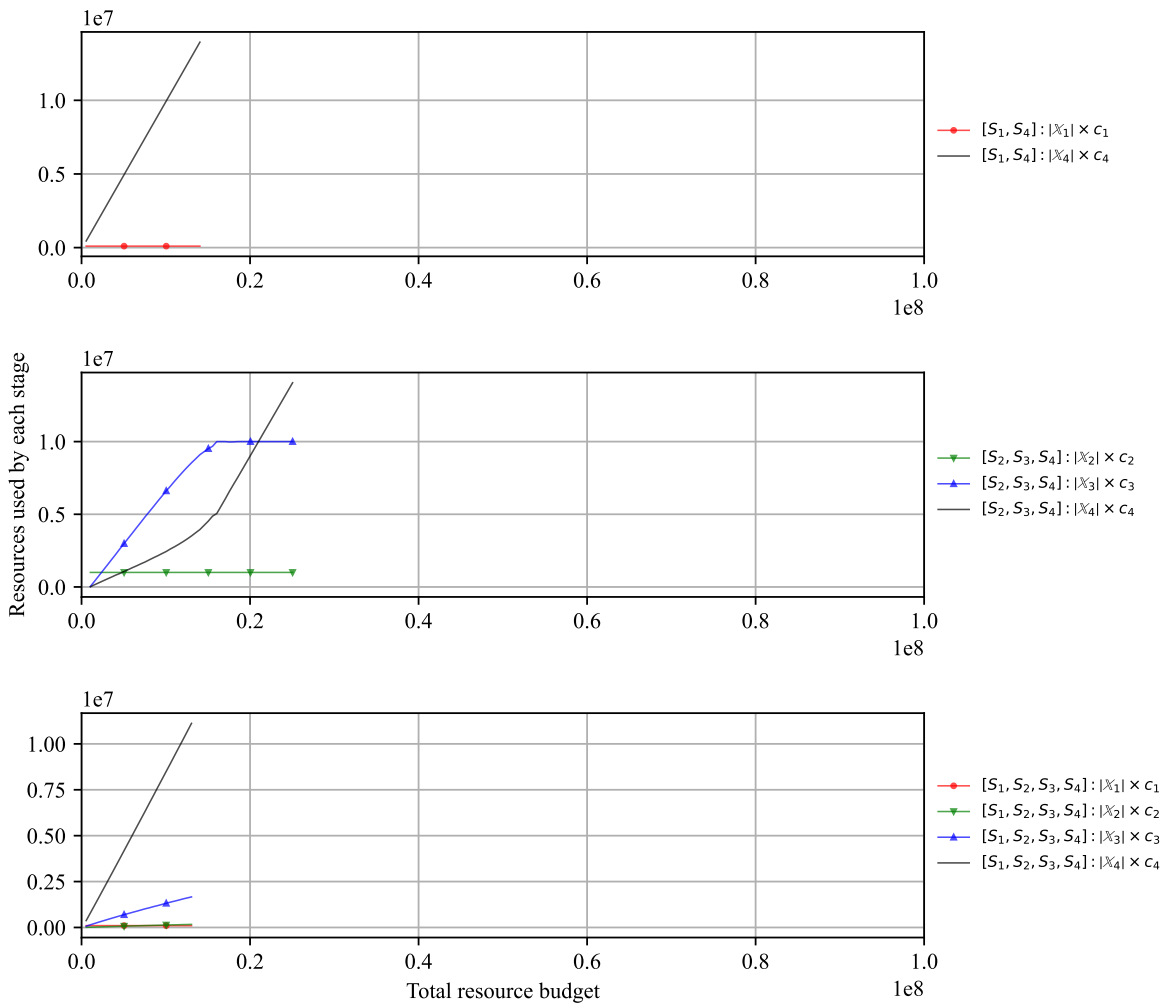


Figure S32: Resources used by each stage.

$$p(y_1, y_2, y_3, y_4) \sim \mathcal{N} \left(\boldsymbol{\mu} = \mathbf{0}, \boldsymbol{\Sigma} = \begin{bmatrix} 1 & 0.8 & 0.2 & 0.8 \\ 0.8 & 1 & 0.2 & 0.8 \\ 0.2 & 0.2 & 1 & 0.2 \\ 0.8 & 0.8 & 0.2 & 1 \end{bmatrix} \right) \quad (9)$$

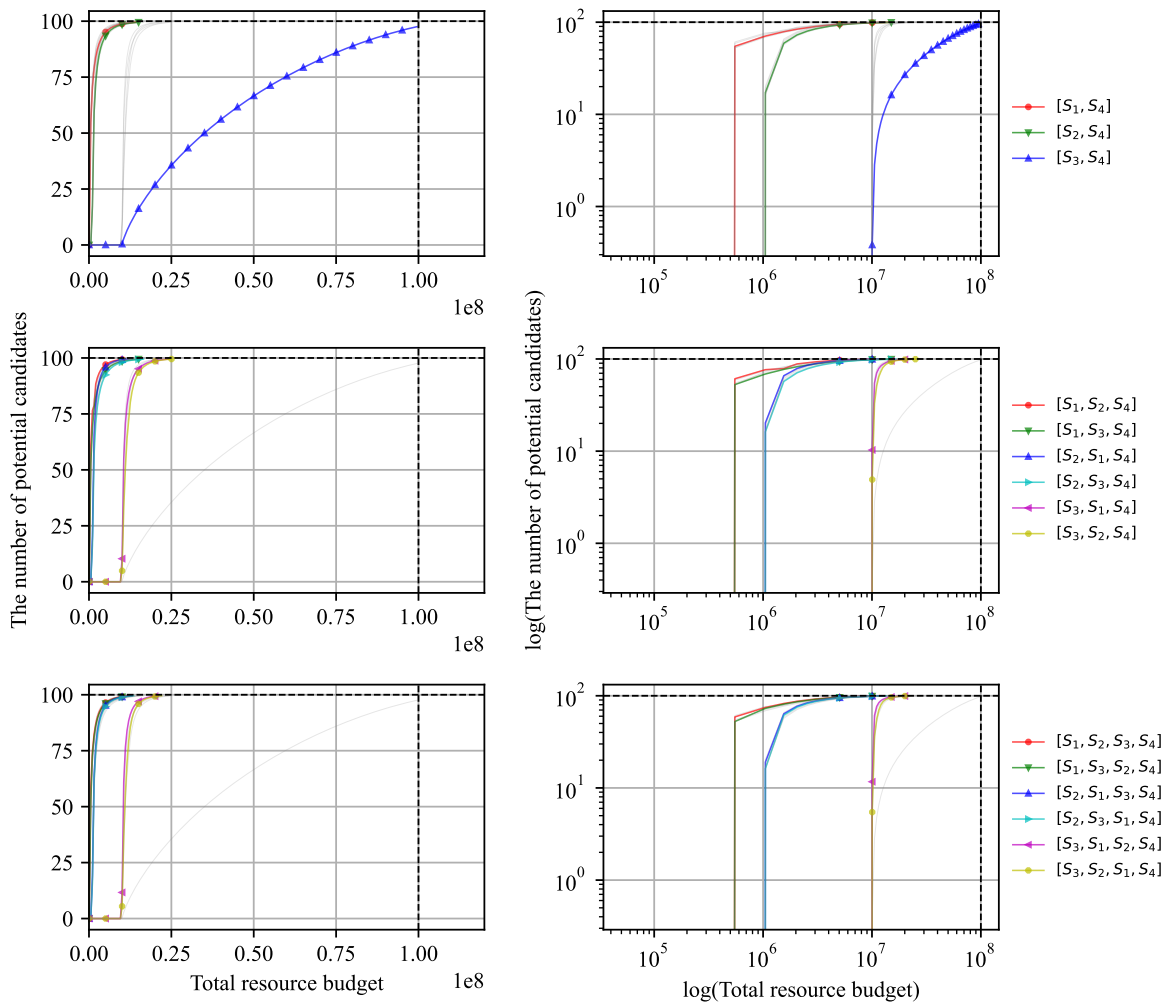


Figure S33: Performance comparison of the optimized high-throughput virtual screening pipelines in terms of discovery capability.

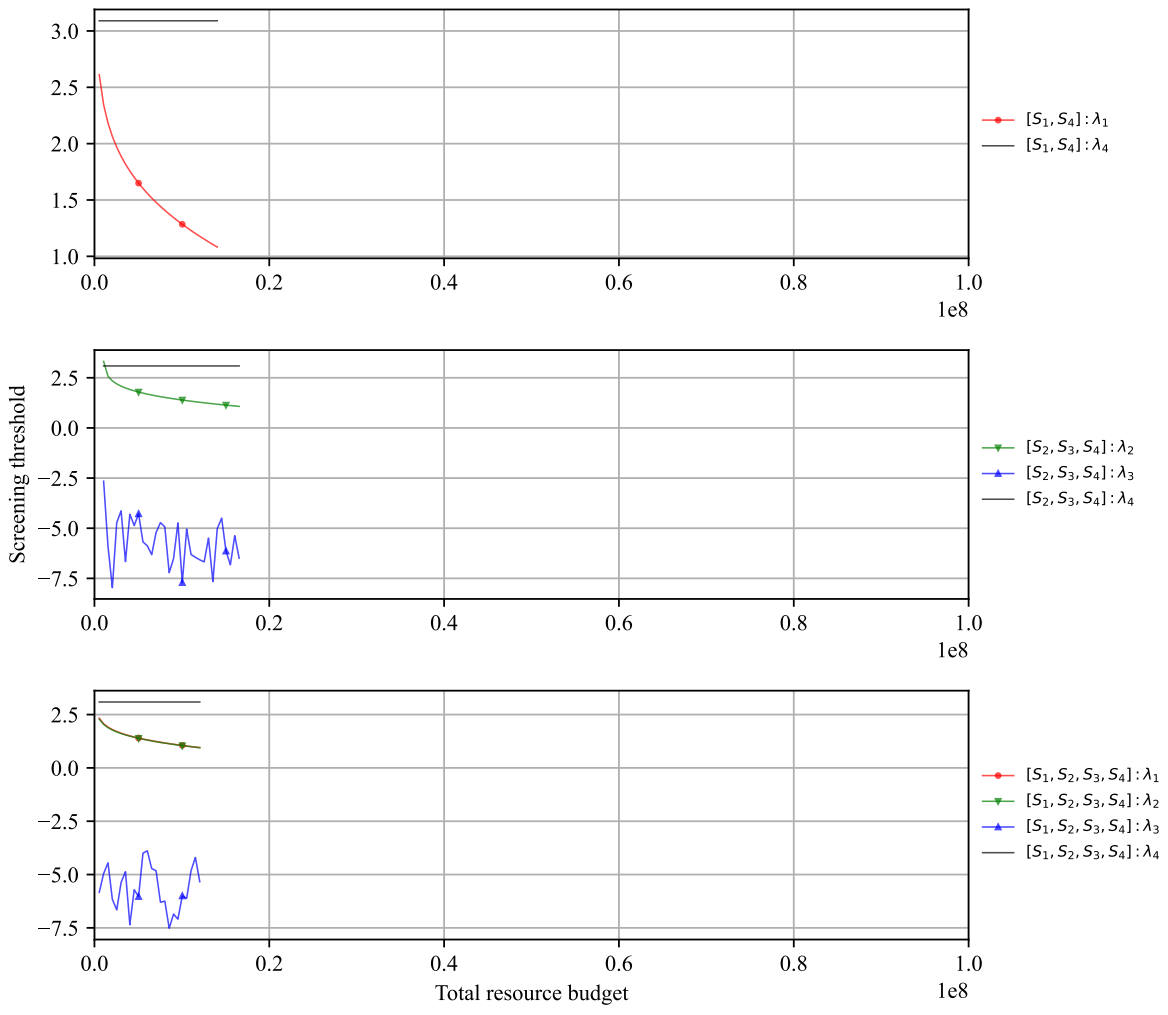


Figure S34: Screening thresholds of the optimized pipelines.

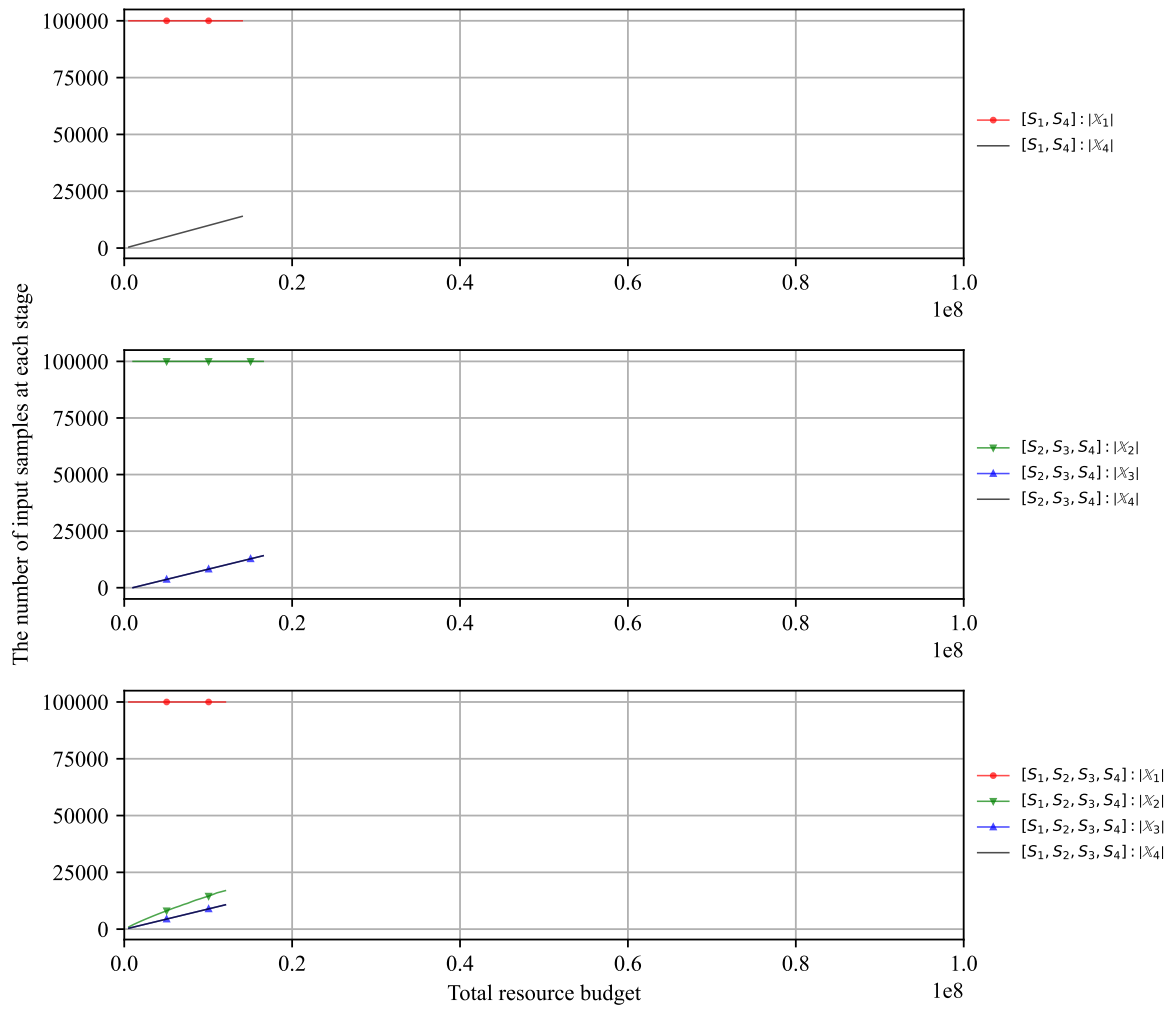


Figure S35: The number of input samples at each stage.

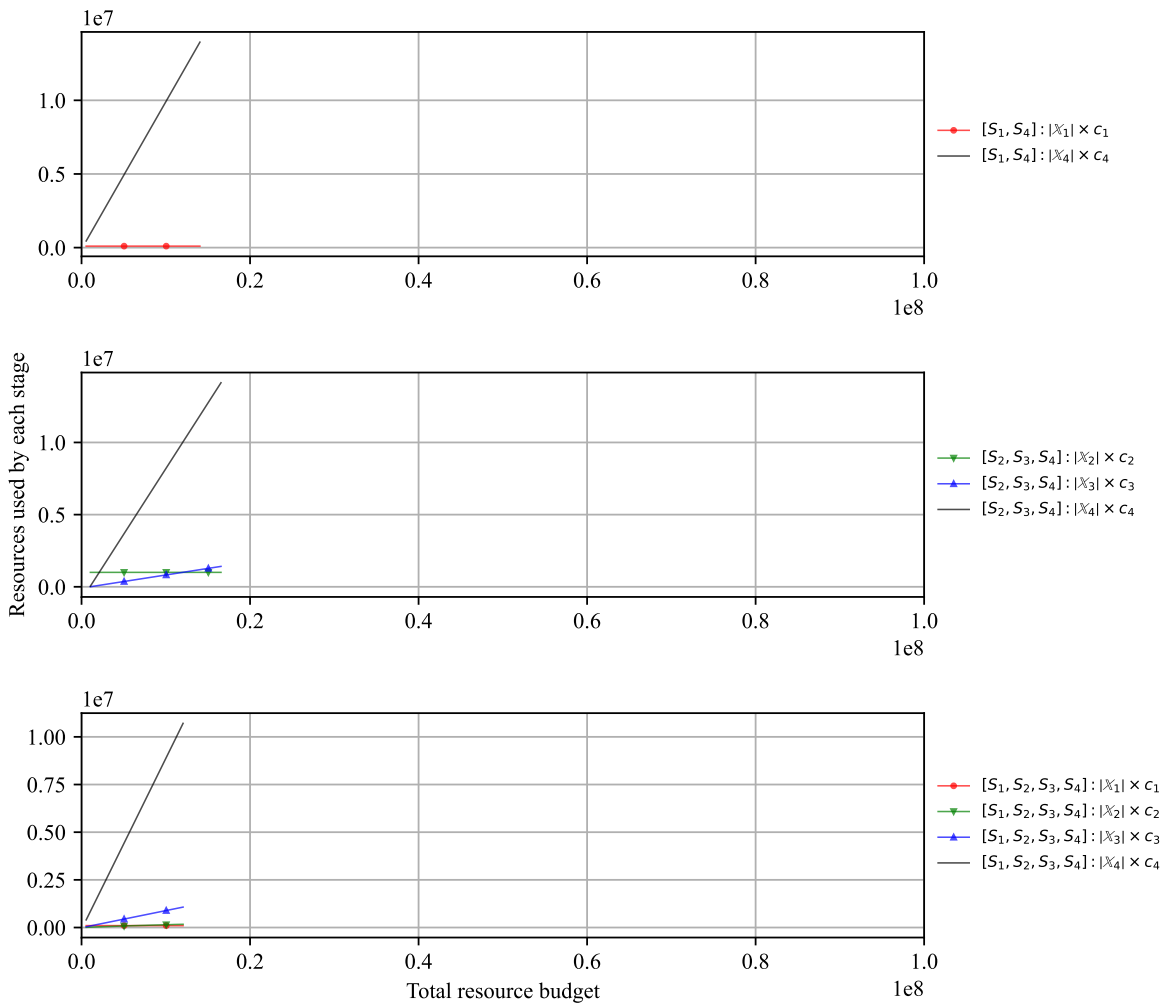


Figure S36: Resources used by each stage.

$$p(y_1, y_2, y_3, y_4) \sim \mathcal{N} \left(\boldsymbol{\mu} = \mathbf{0}, \boldsymbol{\Sigma} = \begin{bmatrix} 1 & 0.3 & 0.2 & 0.1 \\ 0.3 & 1 & 0.3 & 0.2 \\ 0.2 & 0.3 & 1 & 0.3 \\ 0.1 & 0.2 & 0.3 & 1 \end{bmatrix} \right) \quad (10)$$

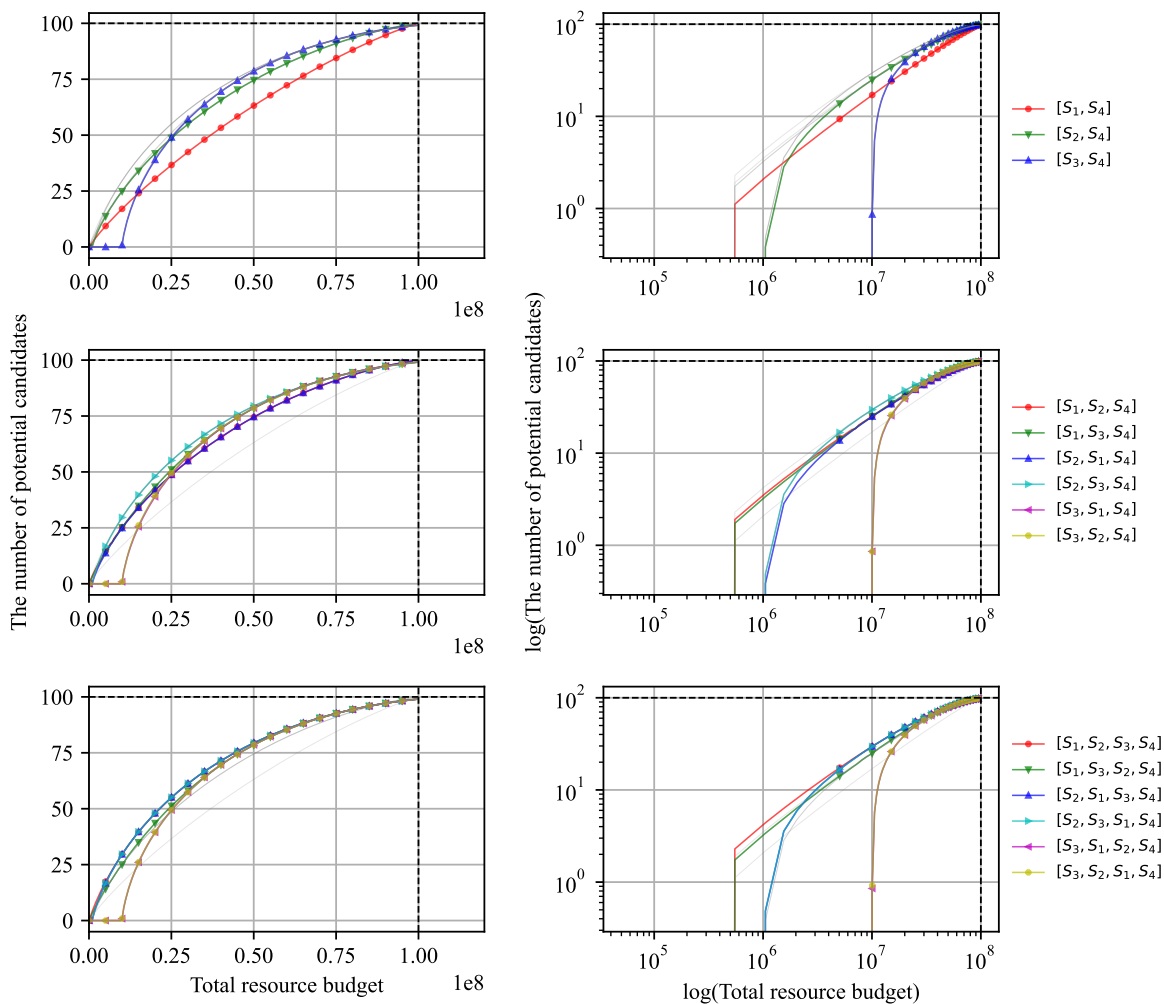


Figure S37: Performance comparison of the optimized high-throughput virtual screening pipelines in terms of discovery capability.

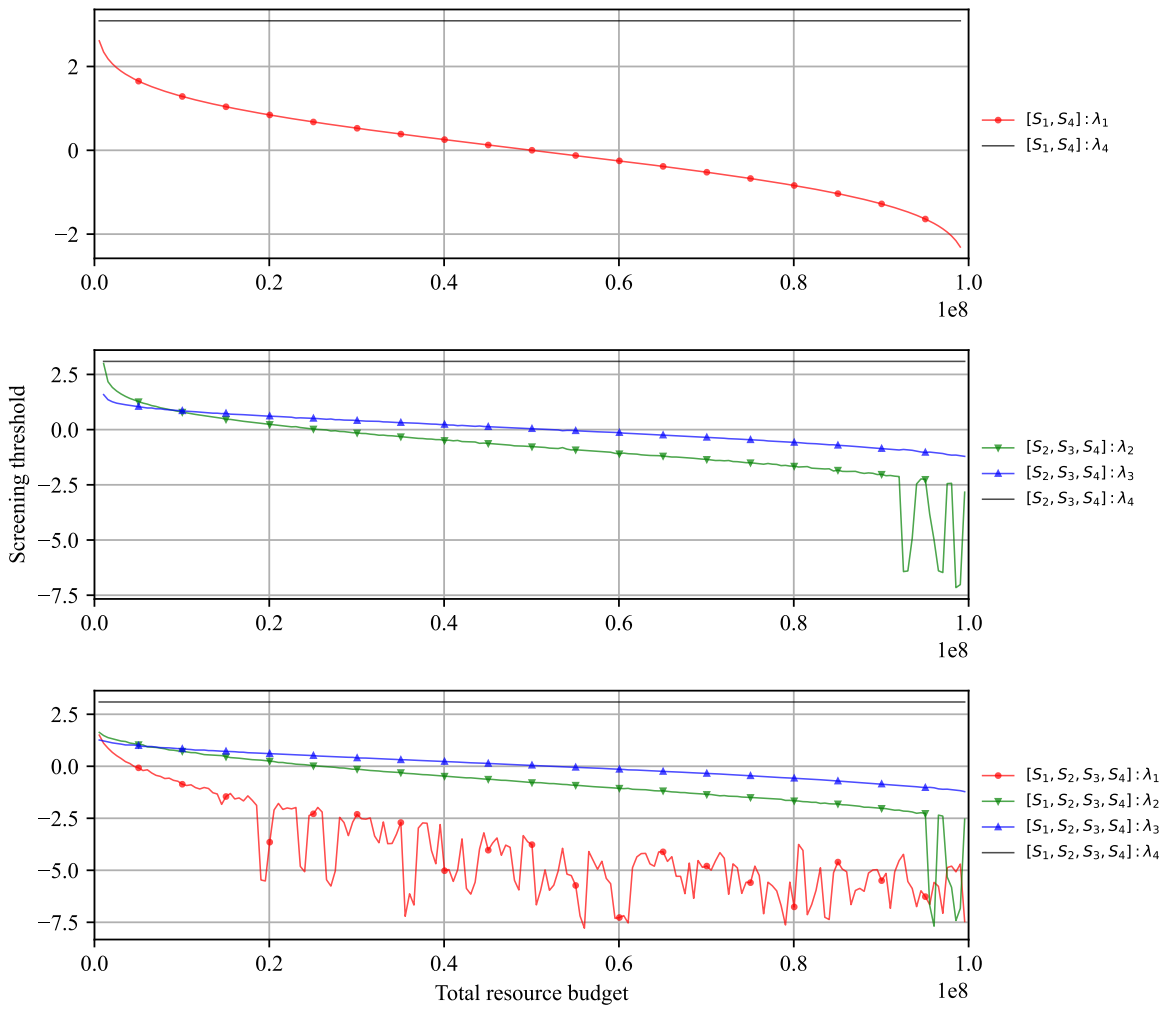


Figure S38: Screening thresholds of the optimized pipelines.

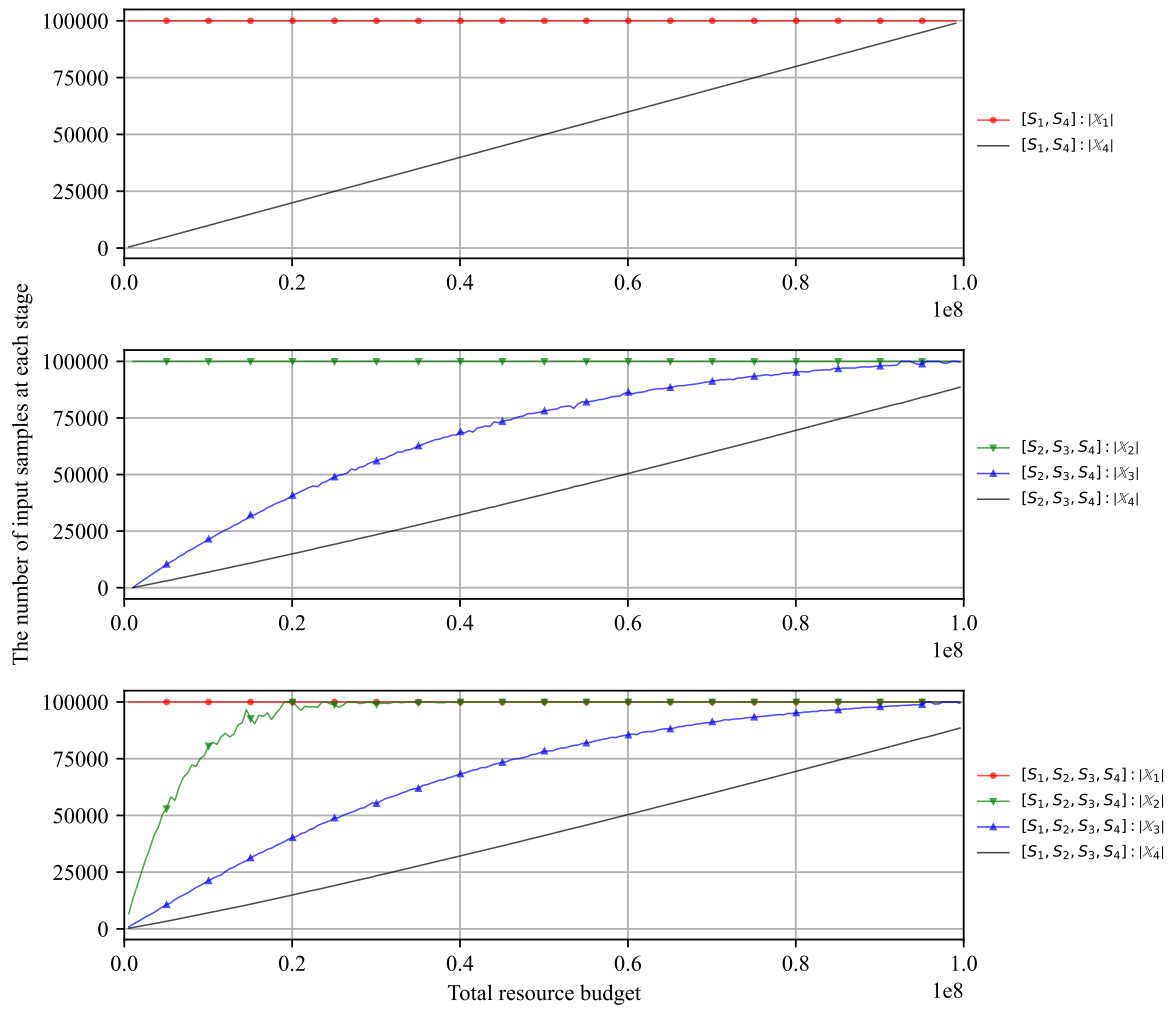


Figure S39: The number of input samples at each stage.

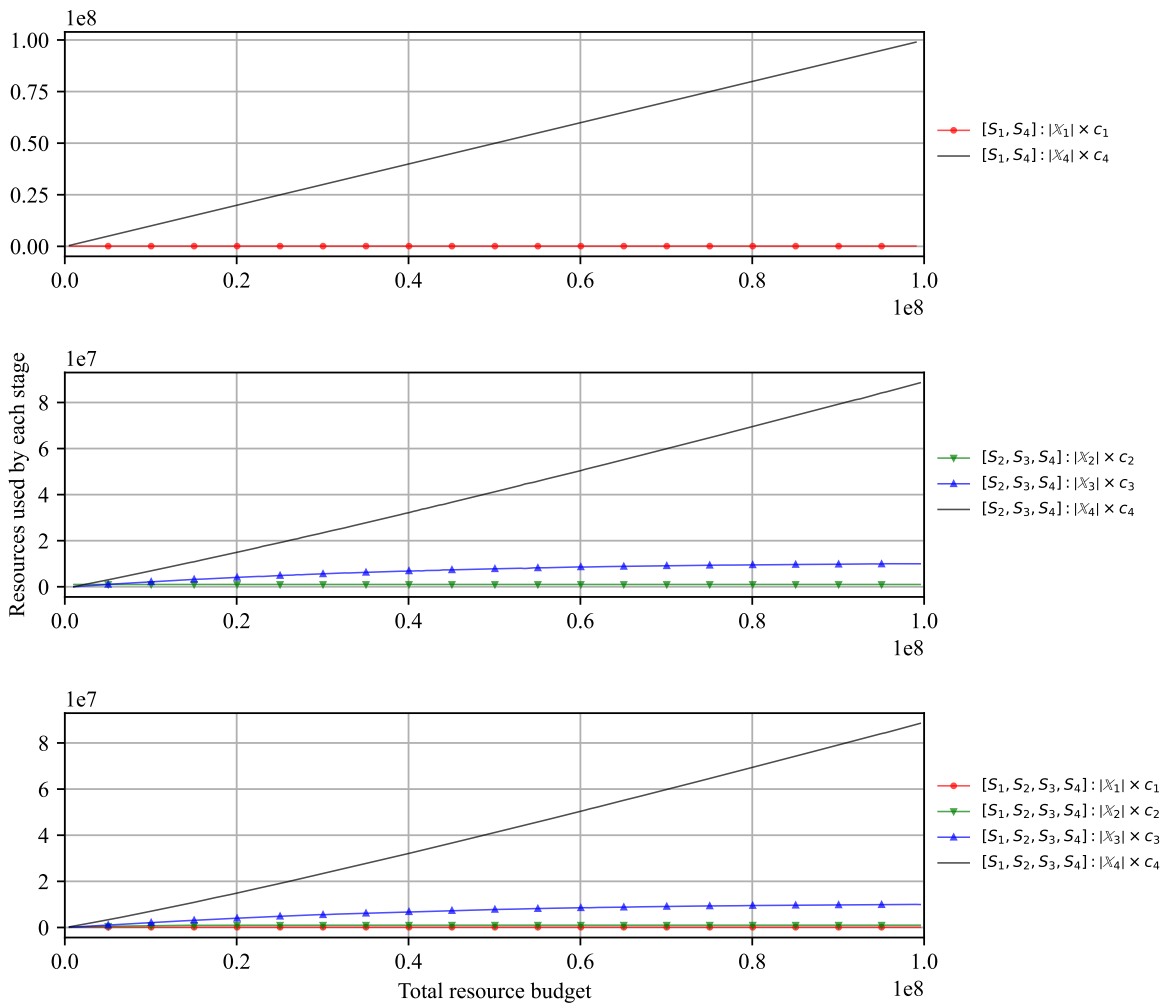


Figure S40: Resources used by each stage.

$$p(y_1, y_2, y_3, y_4) \sim \mathcal{N} \left(\boldsymbol{\mu} = \mathbf{0}, \boldsymbol{\Sigma} = \begin{bmatrix} 1 & 0.8 & 0.7 & 0.6 \\ 0.8 & 1 & 0.8 & 0.7 \\ 0.7 & 0.8 & 1 & 0.8 \\ 0.6 & 0.7 & 0.8 & 1 \end{bmatrix} \right) \quad (11)$$

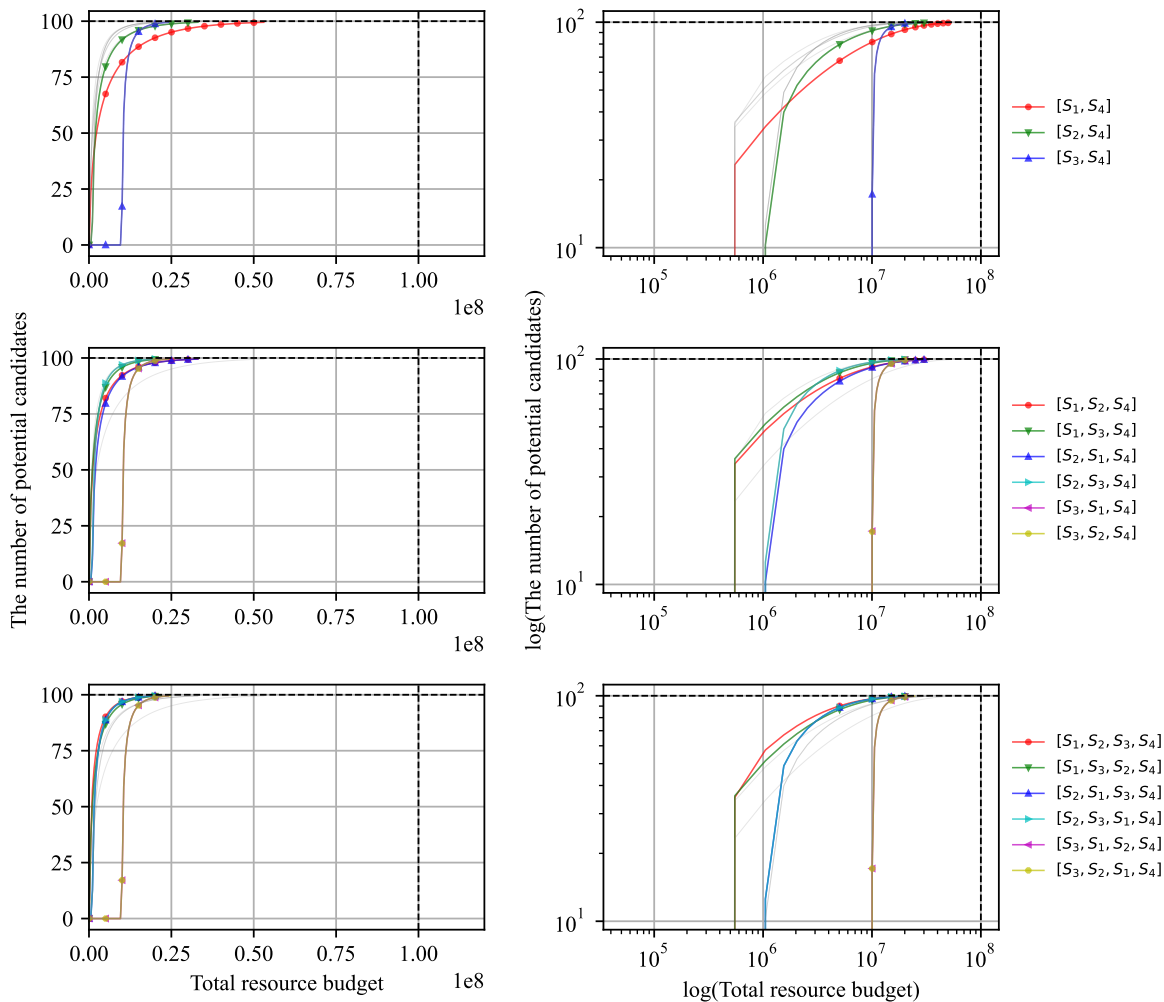


Figure S41: Performance comparison of the optimized high-throughput virtual screening pipelines in terms of discovery capability.

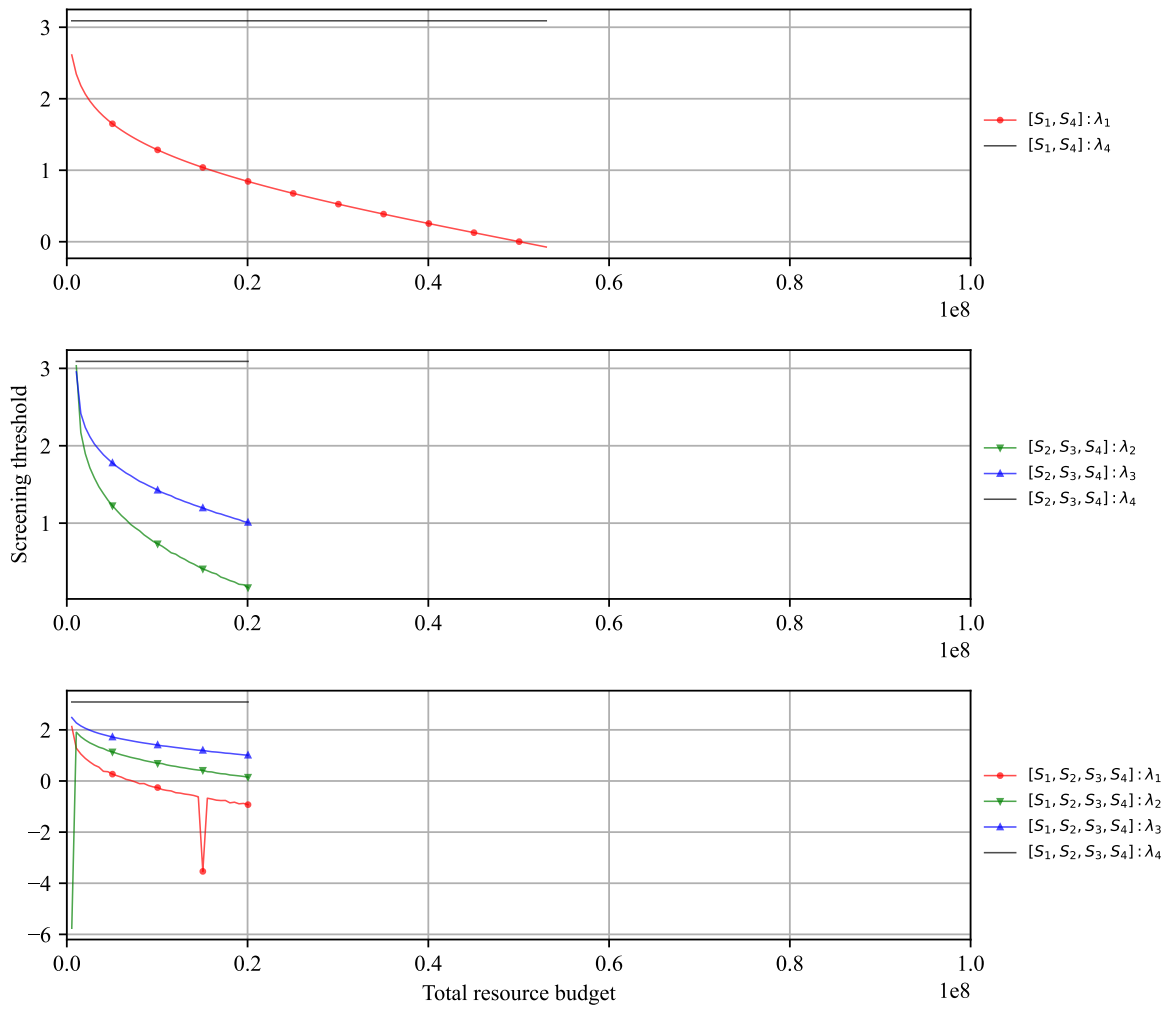


Figure S42: Screening thresholds of the optimized pipelines.

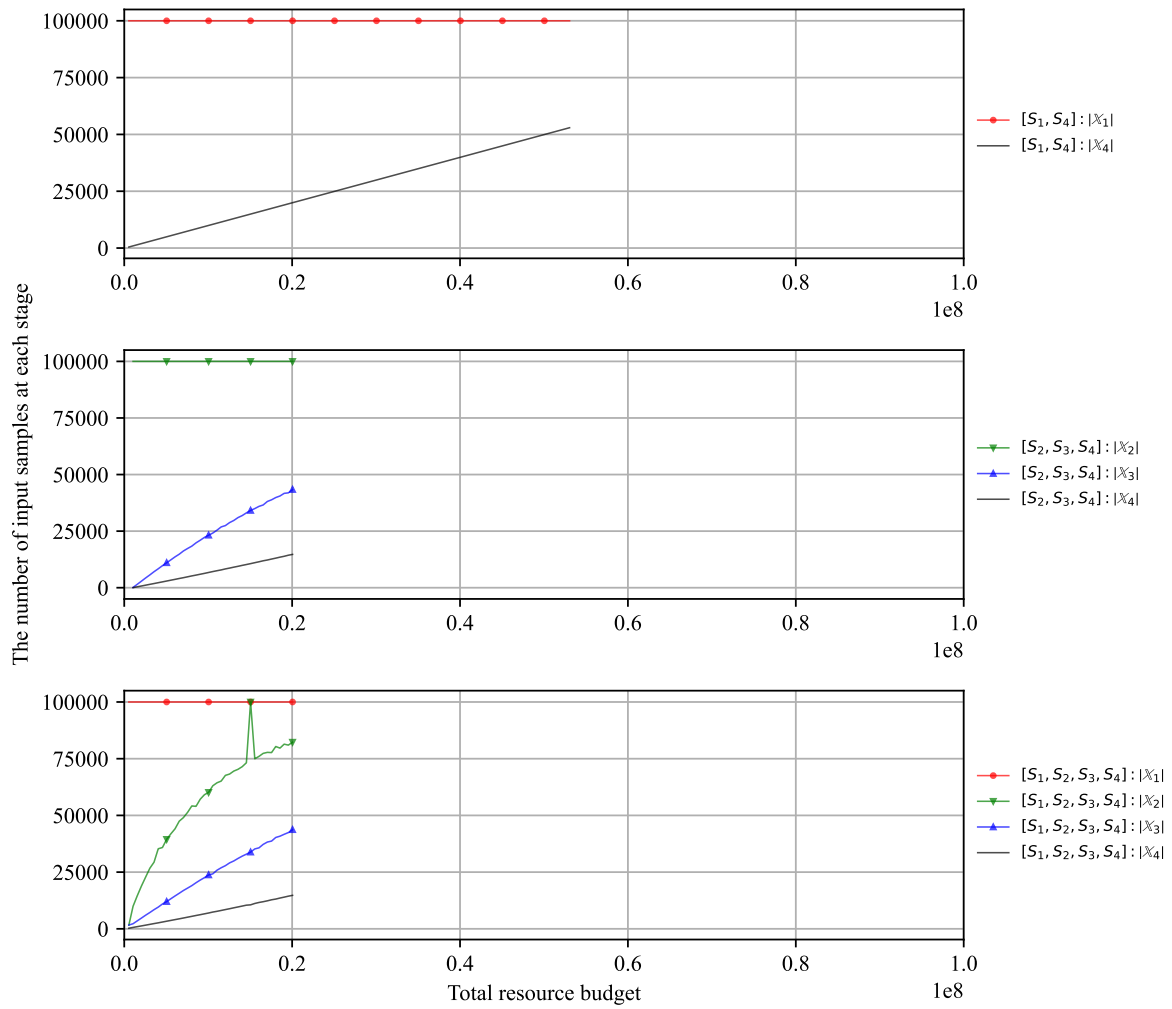


Figure S43: The number of input samples at each stage.

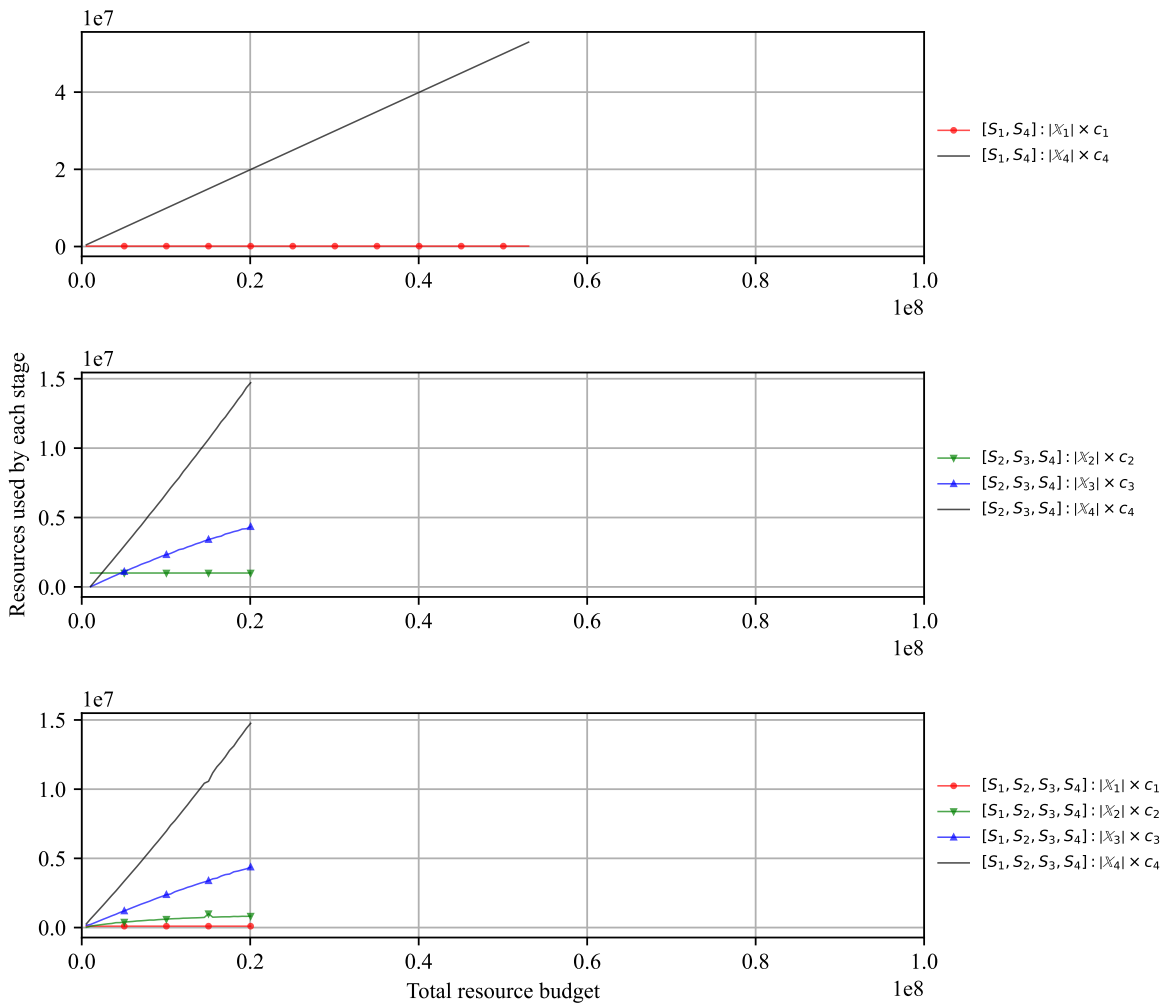


Figure S44: Resources used by each stage.

Performance evaluations of high-throughput virtual screening pipelines for identifying long non-coding RNAs

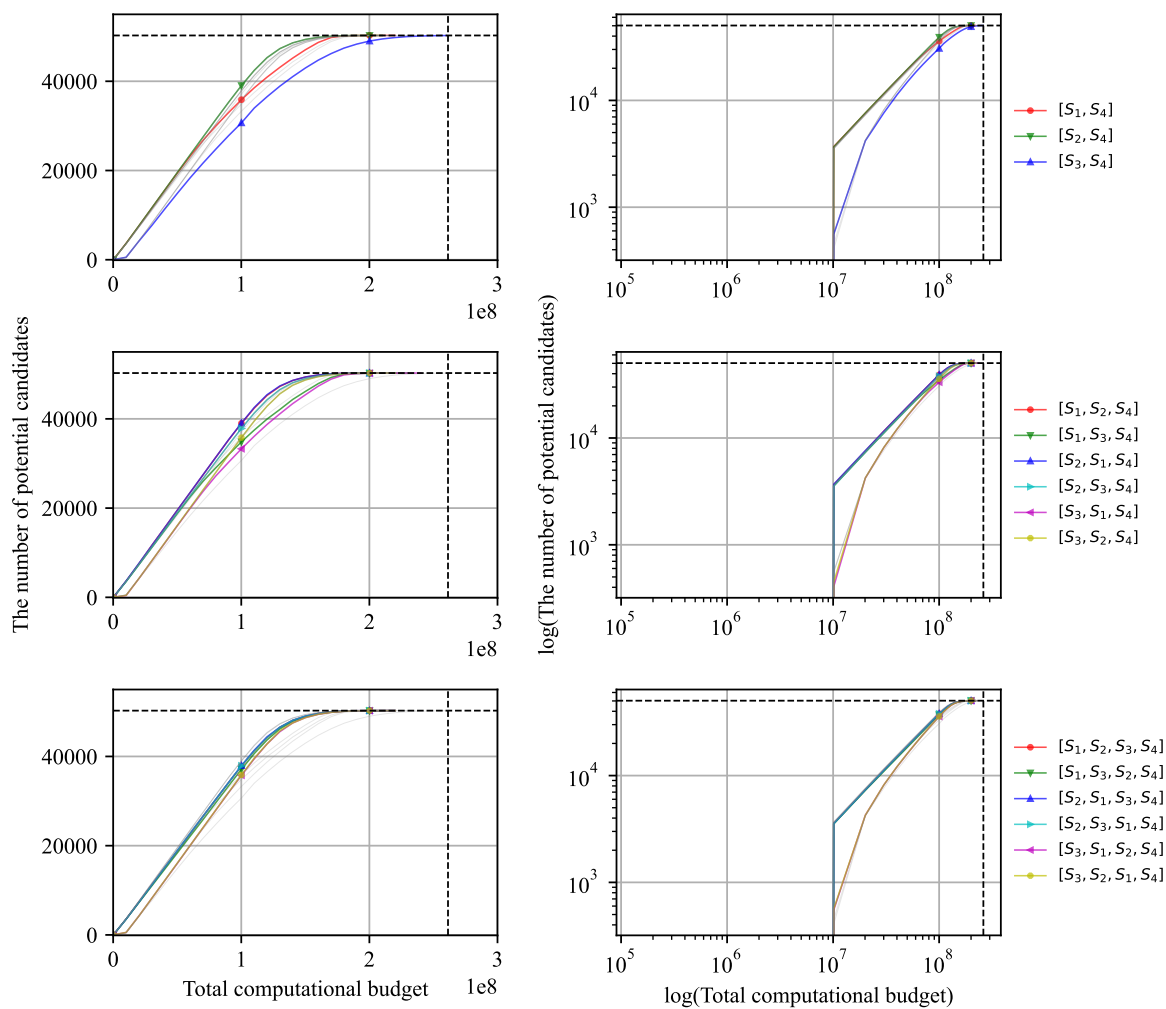


Figure S45: Performance comparison of the optimized high-throughput virtual screening pipelines in terms of discovery capability.

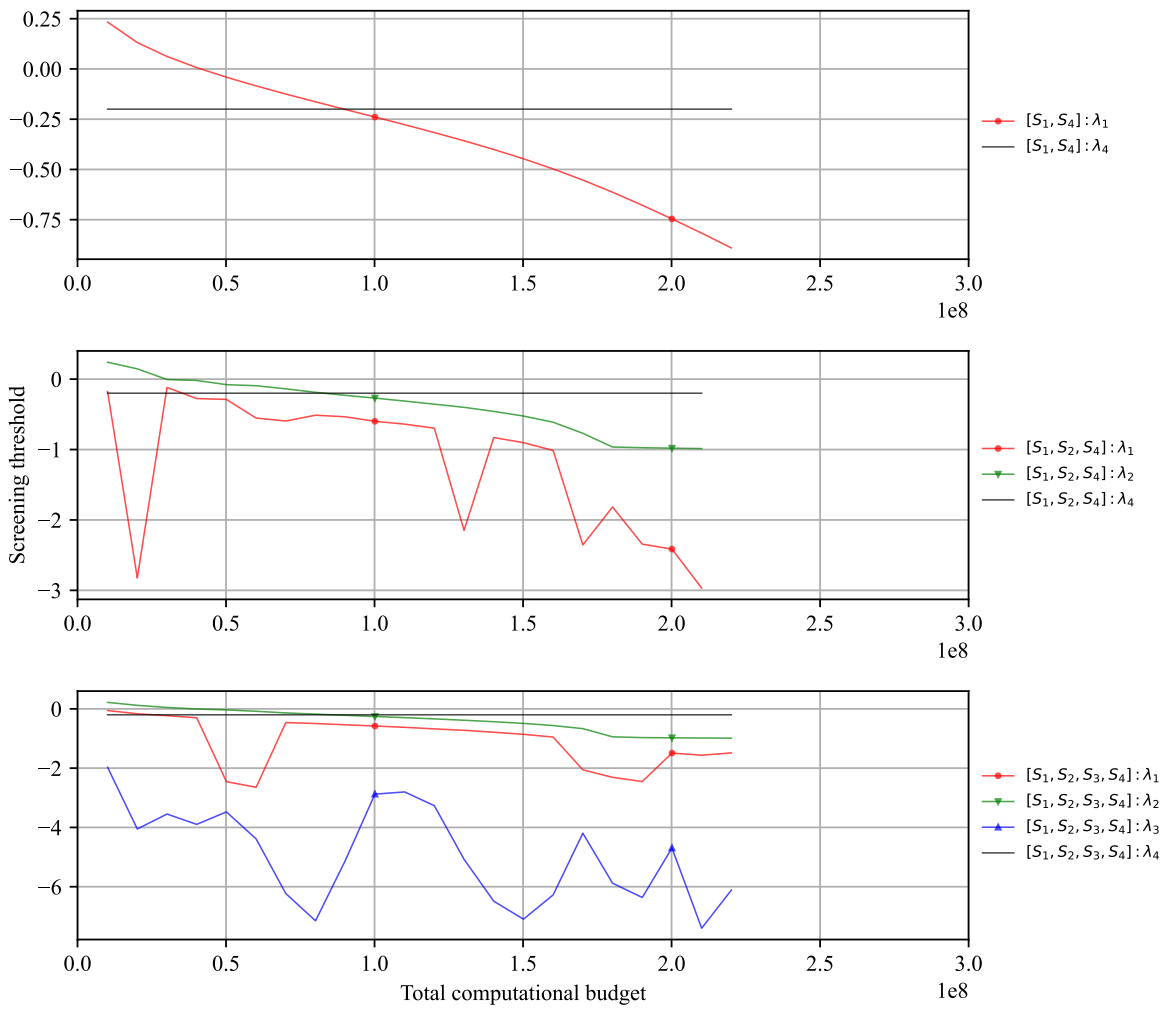


Figure S46: Screening thresholds of the optimized pipelines.

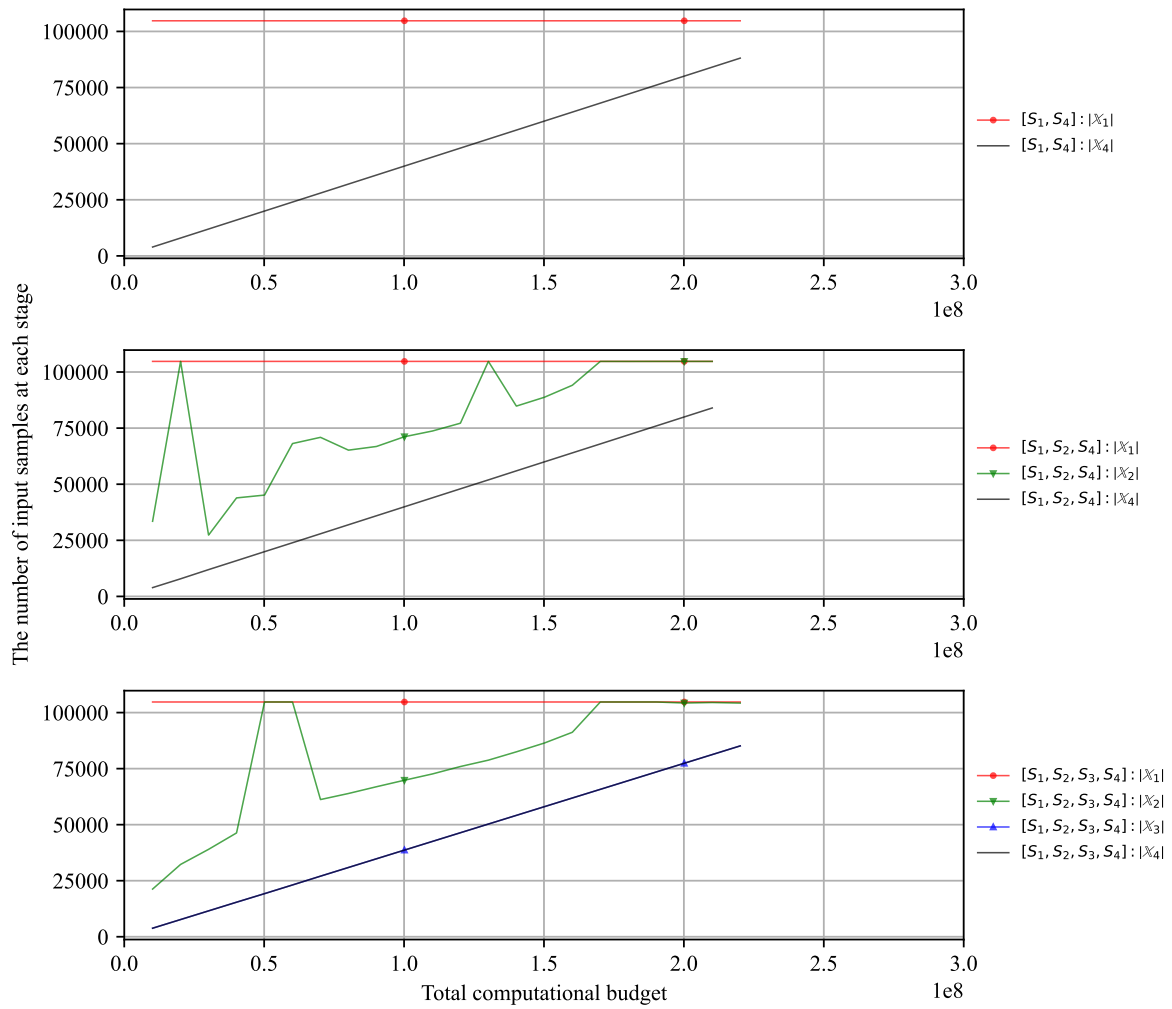


Figure S47: The number of input samples at each stage.

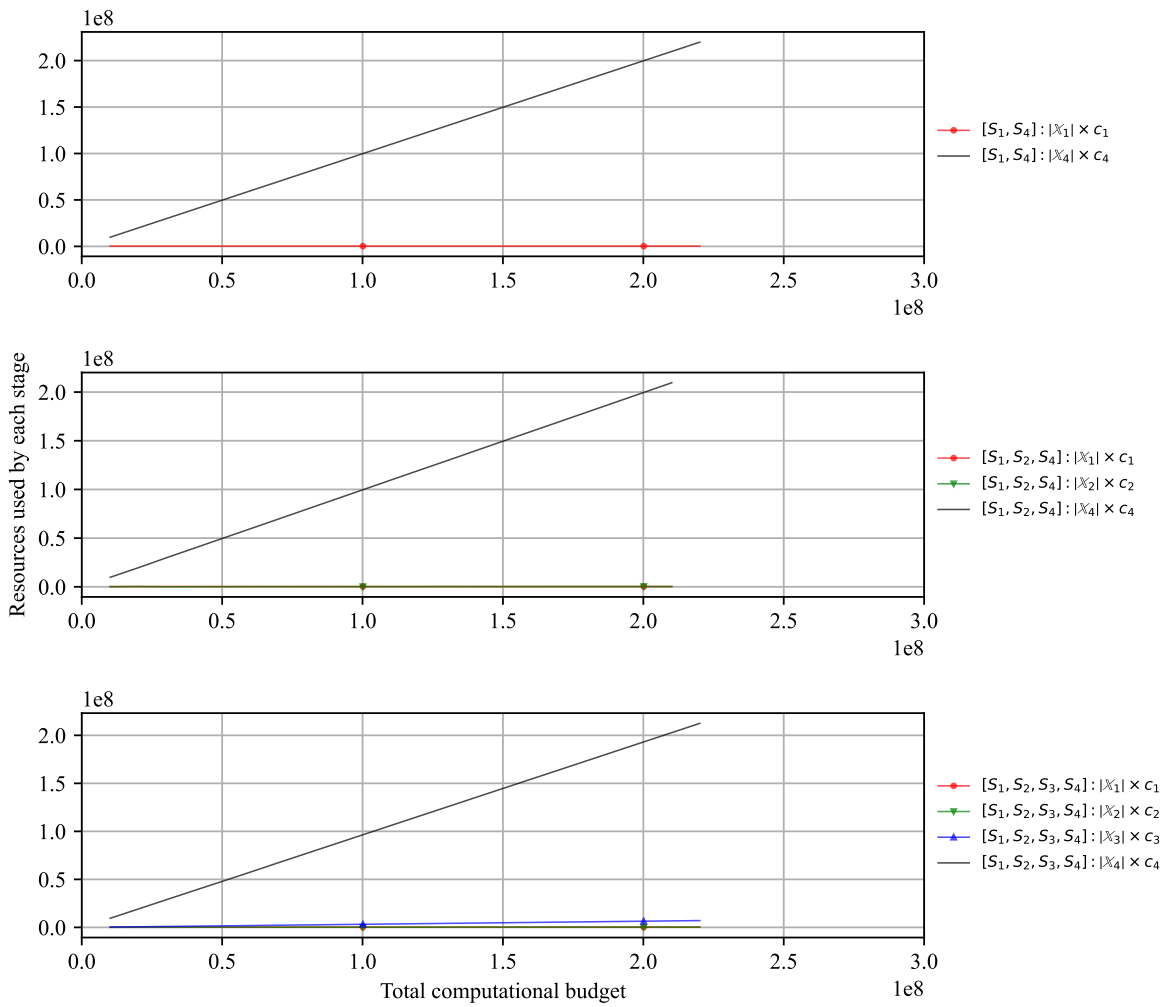


Figure S48: Resources used by each stage.