Supporting Information

Computational Prediction of ω-Transaminase Specificity by a Combination of Docking and Molecular Dynamics Simulations

Carlos Ramírez-Palacios[1,2], Hein J. Wijma[1], Sebastian Thallmair[2,3], Siewert J. Marrink[2], Dick B. Janssen[1*]

[1]Biotransformation and Biocatalysis, Groningen Biomolecular Sciences and Biotechnology Institute (GBB), University of Groningen, Nijenborgh 4, 9747 AG Groningen, Groningen, The Netherlands
[2]Molecular Dynamics, Groningen Biomolecular Sciences and Biotechnology Institute (GBB), University of Groningen, Nijenborgh 7, 9747 AG Groningen, Groningen, The Netherlands
[3]Frankfurt Institute for Advanced Studies, Ruth-Moufang-Str. 1, 60438 Frankfurt am Main, Germany

* Corresponding author
Groningen Biomolecular Sciences and Biotechnology Institute
University of Groningen
Nijenborgh 4
9747 AG Groningen
The Netherlands
E-mail: d.b.janssen@rug.nl

ΔΔG‡ deviations from experimental values

Equation S1 converts the *ee*% for product obtained in an asymmetric synthesis reaction to a fraction of products [S]/[R]:

$$\frac{[S]}{[R]} = \frac{\frac{ee\%}{100} + 1}{1 - \frac{ee\%}{100}} \quad (S1)$$

The relative reaction rates for the amination of the prochiral ketone to form the (*S*)- or (*R*)-enantiomer determine the outcome:

$$\frac{[S]}{[R]} = \frac{v_S}{v_R} \quad (S2)$$

We can define ΔΔG‡ as the difference in the transition state energy for each process:

$$\Delta\Delta G^{\ddagger} = \Delta G_R^{\ddagger} - \Delta G_S^{\ddagger} \quad (S3)$$

In asymmetric transformations, the relation between $\Delta G_R^{\ddagger}$ and $\Delta G_S^{\ddagger}$ is given by:

$$\Delta\Delta G^{\ddagger} = -RTln\left(\frac{v_S}{v_R}\right) \quad (S4)$$

where, R = 8.314 $J\cdot mol^{-1}\cdot K^{-1}$, and T = 300 K.

Therefore, the deviation from the experimental data provided in Table 1 is computed as follows:

$$deviation = \Delta\Delta G_{calc}^{\ddagger} - \Delta\Delta G_{exp}^{\ddagger} = -RTln\left(\frac{v_S}{v_R}\right)_{calc} - -RTln\left(\frac{v_S}{v_R}\right)_{exp} \quad (S5)$$

Additional figures and tables

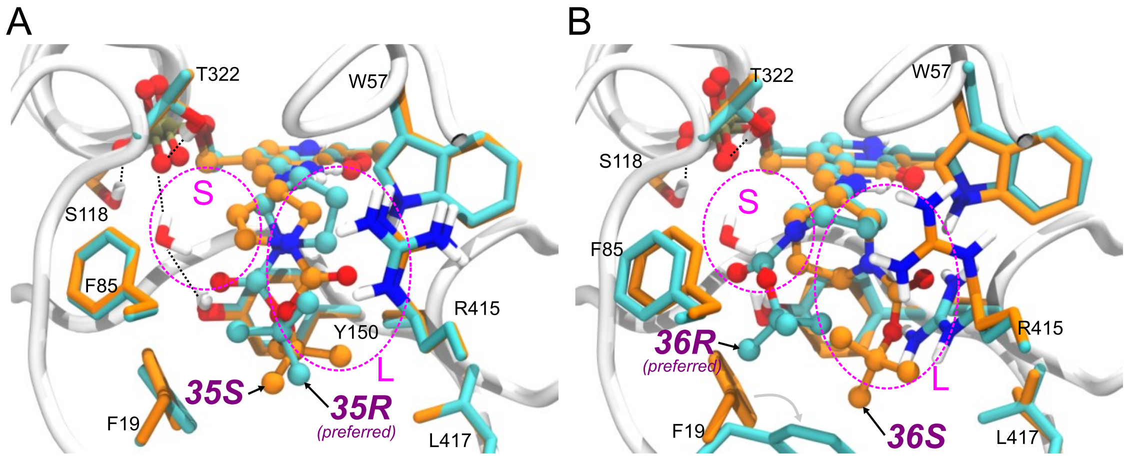

Figure S1. Docked models of the ligands **A) 35R** and **35R**, and **B) 36R** and **36S**. The carbon atoms of the (*S*)- and (*R*)-enantiomer are colored orange and cyan, respectively. The approximate location of the small (S) and large (L) binding pockets are circled in magenta.

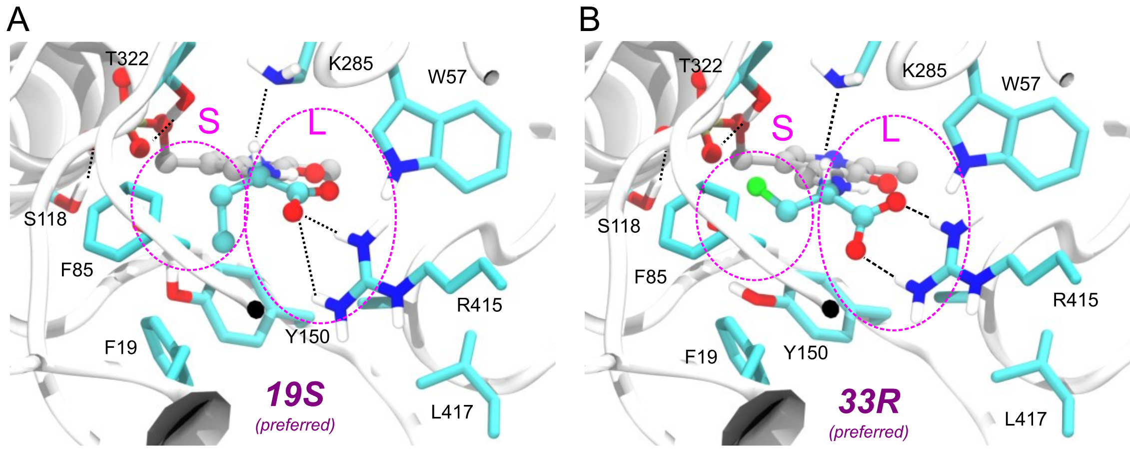

Figure S2. Docked models of external aldimine intermediates of amines **A) 19** (alanine) and **B) 33** (3-fluoroalanine). Both ligands can form either one or two Arg415::COO^{-} salt bridges, which is reflected in the high enantiopreference of *Vf*-TA to produce the (*S*)-amine of **19** and the (*R*)-amine of **33** ($ee\%_{exp}$ is 99 and –98%, respectively). The predictions agree with the experimental observations only when the carboxylate group is modelled in the deprotonated form, otherwise there is no salt bridge formation and the $ee\%_{calc}$ diverges from expected values.

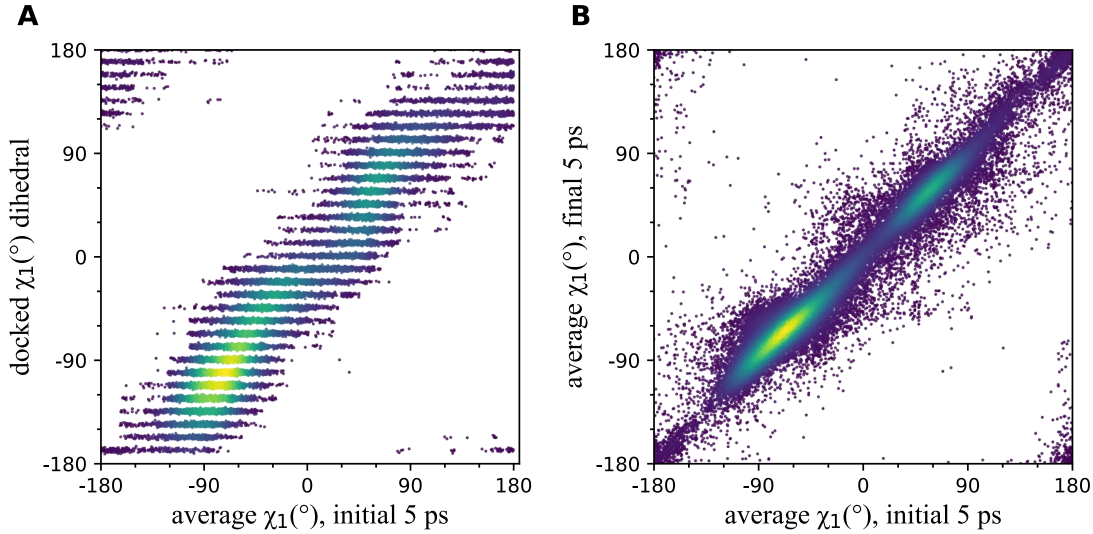

Figure S3. KDE scatter plots showing the variation in the χ_1 dihedral at the beginning and end of the 20 ps MD trajectories. The χ_1 dihedral angle was set to a fixed value at the docking stage but allowed to move freely during the MD simulations. **A)** The dihedral χ_1 did not substantially change its value during the initial 5.0 ps of simulation time respect to the initial value set at the docking stage. **B)** Further, no substantial difference is observed between the value of the χ_1 dihedral during the final 5.0 ps of simulation time with respect to the initial 5.0 ps. Both plots were made using data from all simulations (49 compounds, 2 enantiomers per compound, 64 docked structures per ligand, 5 seeds per docked structure = 31,360 simulations). The color of each dot corresponds to the density of observations in that area, with bluish purple corresponding to the lowest density.

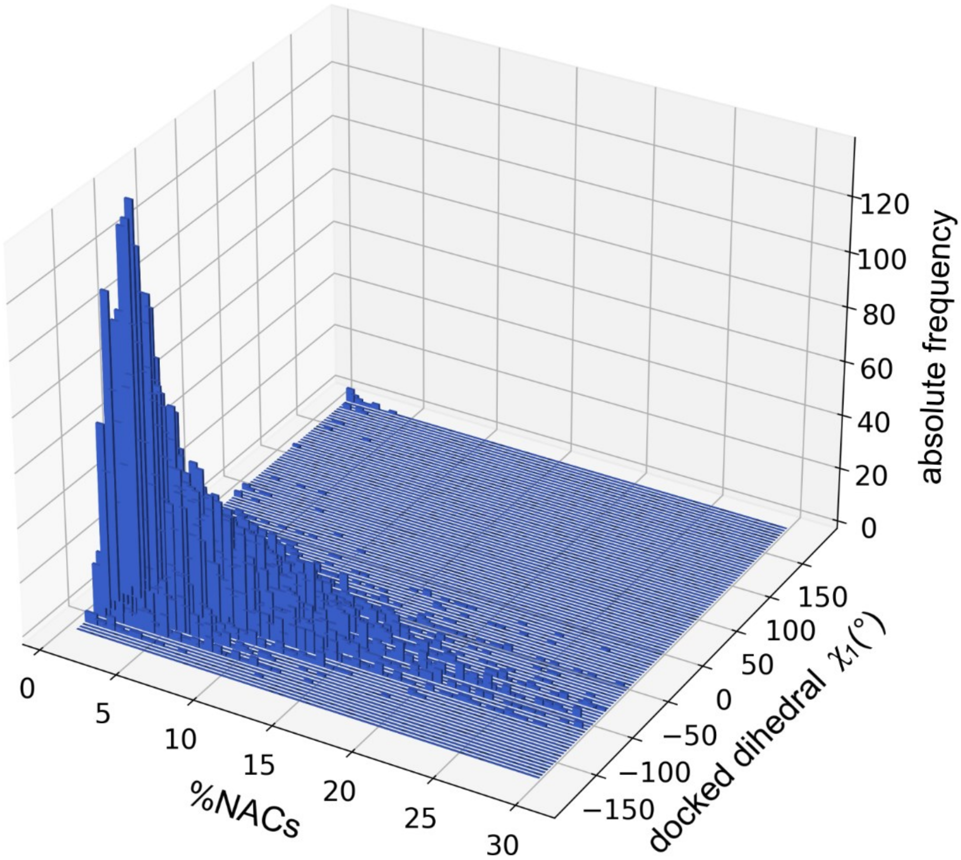

Figure S4. 3D histogram showing the distribution of trajectories with calculated %NACs as a function of their initial χ_1 dihedrals. The histogram includes the NAC occurrence of all trajectories from all compounds in the benchmark dataset (49 compounds, 2 enantiomers per compound, 64 simulations per enantiomer, 5 seeds per simulation). NAC occurrence during the MD trajectory is strongly influenced by the initial χ_1 dihedral. The closer the docked external aldimine complexes were to a catalytic orientation (χ_1 = –90º), the higher chance of producing NACs during the 20 ps MD trajectory. The 20 ps of simulation time is not long enough to allow the χ_1 dihedral to evolve away from its initial configuration.

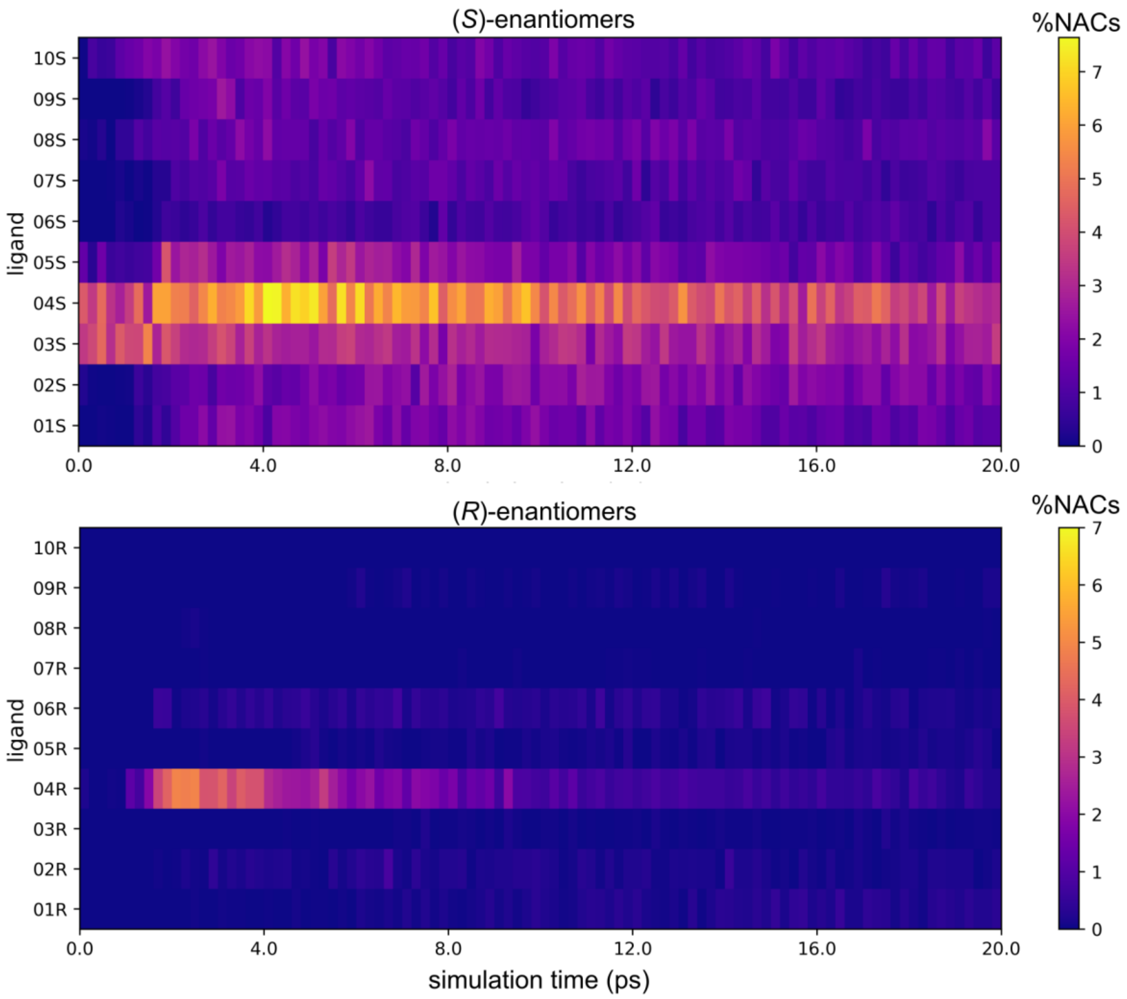

Figure S5. NAC occurrence measured at different simulation windows. The frequency of NAC occurrence was measured using a sliding window (size = 0.2 ps) along the MD trajectory. The entire dataset consisted of 49 unique compounds, but only compounds **01 – 10** are shown in this figure. (*S*)-and (*R*)-enantiomers are shown in the *top* and *bottom* panels, respectively. Because the NAC occurrence is evenly distributed along most of the trajectories, $ee\%_{calc}$ is not dependent on simulation length.

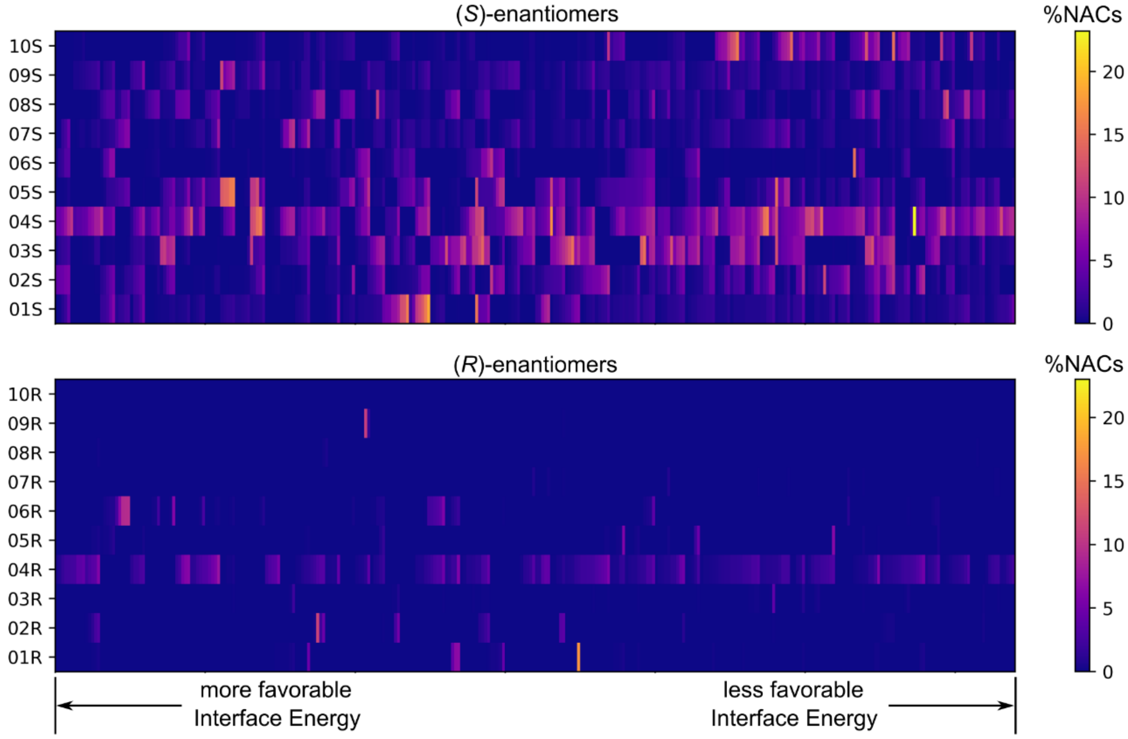

Figure S6. Heatmap showing the NAC occurrence obtained from MD simulations of several ligands (*y*-axis) sorted by the Rosetta score of the starting structures (*x*-axis). Within the set of selected structures (top20%), the NAC-producing potential of a protein-ligand complex is independent from its Rosetta Interface Energy. For any given compound, the frequency of NACs observed in the 64×5 MD simulations (64 docked structures per compound; 5 seeds per docked structure) is shown in the heatmap. The docked structures were sorted along the *x*-axis from more favourable (*left*) to less favourable (*right*) Rosetta Interface Energy. The entire dataset consisted of 49 unique compounds, but only compounds **01** – **10** are shown for clarity. (*S*)-and (*R*)-enantiomers are shown in the *top* and *bottom* panels, respectively. Because the NAC-producing potential of a docked structure is not dependent on its docking score, the $ee\%_{calc}$ is mostly unaffected by whether the top10% or top20% starting structures are selected for MD simulations.

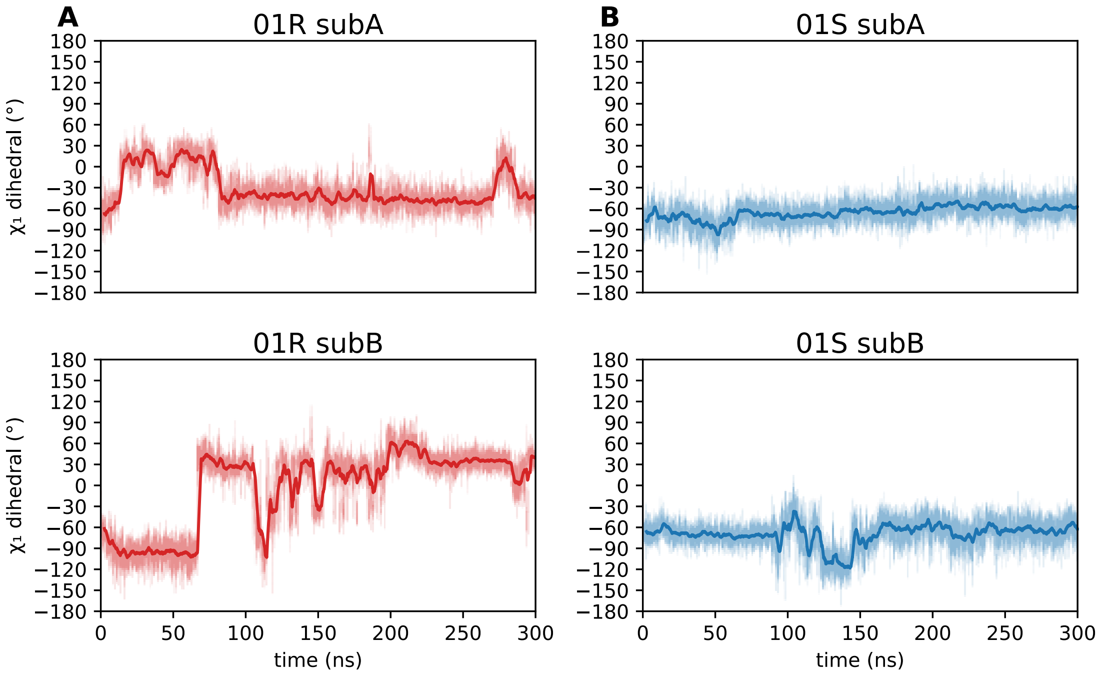

Figure S7. Time evolution of the χ_1 dihedral in 300 ns MD simulations. A) Ligand **01R** bound in the binding site mainly formed by residues of subunit A (*top*) and subunit B (*bottom*). **B)** Ligand **01S** bound in the binding site mainly formed by residues of subunit A (*top*) and subunit B (*bottom*). Ligand **01R** means "the external aldimine of compound **01** that would lead to the production of the (*R*)-amine". The χ_1 dihedral of the docked complex was –90.0° in all cases. The figure shows that the χ_1 dihedral evolves slowly: in the simulation of **01R** (subunit B) it takes 65 ns for χ_1 to jump out of the catalytic region ($\chi_1 = -90 \pm 15°$) to a new region ($\chi_1 \approx +30°$), where it remains for 40 ns. Therefore, short MD simulations (tens of ps) would be unable to sample χ_1.

Table S1. Comparison of the *ee*% calculated using additional simulation setups. MD simulations were run for either 20 ps with 5 random initiation seeds (5 × 20ps) or 100 ps with only one initiation seed (1 × 100ps). The proportion of Rosetta structures that were selected to be used as starting conformations for MD were either the top20% (64 out of 320 structures) or top10% (32 out of 320 total structures). An additional column is included where no scanning of the χ_1 dihedral was performed by Rosetta, and instead all ligands were docked with $\chi_1 = -90^o$. From this table, the best simulation setup is 5 × 20 ps + top20% (with χ_1 scanning) as the $ee\%_{calc}$ most closely matches the $ee\%_{exp}$.

Compound	$ee\%_{exp}$	$ee\%_{calc}$				
		5×20 ps		**1×100 ps**		**5×20 ps**
		top20%	**top10%**	**top20%**	**top10%**	$\chi_1 = -90^o$
01	99%	82	82	63	48	71
02	84-89%	84	95	88	99	80
03	97-99%	98	97	91	99	12
04	93-99%	56	60	72	75	74
05	92-96%	92	85	81	85	95
06	96-99%	47	85	72	94	-2
07	98-99%	99	97	92	95	-65
08	66-99%	100	100	100	100	-13
09	99%	94	100	61	100	-31
10	96-99%	100	100	100	100	48
11	95-99%	100	100	100	100	10
12	90%	98	99	90	93	83
13	99%	4	55	-25	12	-28
14	80%	99	99	99	99	98
15	99%	-37	-47	-46	-57	-65
16	99%	98	100	92	90	70
17	100%	100	100	100	100	-64
18	90%	84	71	83	73	-4
19	99%	95	100	97	100	30
20	96%	96	100	91	100	-44
21	99%	92	96	73	94	21
22	99%	100	100	100	100	3
23	99%	99	97	99	98	-31
24	98%	97	98	94	95	84
25	99%	100	100	100	100	-18
26	99%	72	100	52	100	-62
27	99%	83	96	79	100	-9
28	81%	49	55	71	79	58
29	99%	100	100	100	100	-79
30	95%	91	98	93	96	77
31	99%	78	89	77	78	-52
32[#]	-98%	34	35	32	29	79
33[#]	-98%	-96	-100	-98	-100	-3
34	40%	-42	-40	-24	-9	-61
35	-99%	42	-23	13	-45	61
36	-98%	9	-7	-6	-29	-70
37	99%	95	99	96	99	53
38	99%	83	91	75	90	33
39	99%	100	100	100	100	22

40	18-54%	-31	-47	-34	-52	-22
41	76%	23	41	-1	18	-22
42	99%	1	6	-5	11	-37
43	97-98%	-11	-36	-33	-59	--
44	93-99%	55	50	23	20	58
45[#]	-1 to -40%	-59	-69	-62	-75	-74
46	99%	68	95	74	-21	99
47	15%	56	50	75	70	1
48	70%	80	97	81	89	22
49	90%	87	87	90	88	79

[#] Different R/S notation due to shift in CIP, but preferred enantiomer with similar stereoconfiguration as (*S*)-**01**.