

**Supporting information for:
Predicting Productive Binding Modes for Substrates and Carbocation Intermediates in
Terpene Synthases – Bornyl Diphosphate Synthase as a Representative Case**

Terrence E. O'Brien,[‡] Steven J. Bertolani,[‡] Yue Zhang,[‡] Justin B. Siegel,^{‡,x,y*} and
Dean J. Tantillo^{‡*}

[‡]Department of Chemistry, University of California, Davis, CA 95616, USA

^xDepartment of Biochemistry and Molecular Medicine, University of California Davis, Davis, California, USA.

^yGenome Center, University of California Davis, Davis, California, USA.

Corresponding Authors

* Justin B. Siegel, jbsiegel@ucdavis.edu

* Dean J. Tantillo, djtantillo@ucdavis.edu

Content

S1: Explanation on the interpretation of the energy scores in Rosetta.

S2: Figure S1: Heatmap from **Figure 2** in the main text with docking scores included.

S3: Table S1. Values of the constraints used in docking the diphosphate complex for all docking runs.

S4: Figure S2. Testing for convergence for docking into BPPS crystal structure.

S5: Figure S3. Heatmap for docking the known stereochemistry structures into alternate crystal structure, 1N23.

S6: Figure S4. RMSD analysis for the 1N23 docking results.

S7: Figure S5. Pathway of least movement identified by the RMSD search for 1N23.

S8: Figure S6. Heatmap for docking the enantiomer stereochemistry of the carbon skeleton into 1N23.

S9: Figure S7&8. Details on alternate filtering for *ent-D* and their results.

S10: Discussion on the formation of (-)-BPP from (-)-LPP using (+)-bornyl diphosphate synthase from sage.

S11: Table S2. Energies of the 1-iodogermanene used in conformational search.

S12: Figure S9. Heatmap which details the effect of the new Rosetta energy function on the BPPS prediction.

S13: References

Explanation on meaningful difference of scores.

There is an inherent noise in any calculation, where approximations used to simplify the math may impact the accuracy of said calculation. This is even more true for force-field based calculations, which often greatly simplify the math in exchange for much quicker calculations. This can lead to situation where, what may look like a difference in some scoring metric is, in fact, not actually a difference. The rule of thumb we use is that score differences of less than 3 Rosetta energy units are effectively indistinguishable. For example, if there was a docking score for one pose of a ligand that was -4.5 Rosetta energy units and another was -6.0 Rosetta energy units, that is only a difference of 1.5 Rosetta energy units and is effectively the same score. When comparing scores (and their resulting poses), we would like to see a difference of ~5 Rosetta energy units in order to make a confident prediction about which pose is more likely to be relevant. To be clear, this isn't a Rosetta problem, but a reality of the approximations used in many docking simulation programs.

In the initial development of *TerDockin*, we had hoped that Interface Energy (the Rosetta equivalent to docking score) might discriminate a "good" pose from a "bad" one. This isn't the case (See **Figure S1**). For every intermediate in any given orientation the Interface Energy was equivalent, given the above mentioned noise in these calculations. The average interface energy for orientation 1, intermediate **C** is -5.95 and for orientation 5 is -5.70, illustrating the challenge in using docking score to make a prediction. In retrospect, this makes sense, as it is unsurprising that a force-field-based approach is insensitive to some of the subtleties of interactions which govern binding in this family of enzymes.

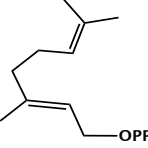
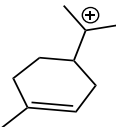
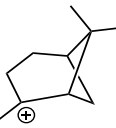
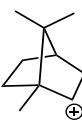
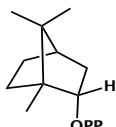
		Orientation							
		1	2	3	4	5	6	7	8
Intermediate		★	-	-	-	-	-	-	★
		GPP attached to oxygen A				GPP attached to oxygen B			
GPP		88 <i>381</i> <u>-8.62</u>				12% <i>52</i> <u>-8.49</u>			
A		100% <i>456</i> <u>-7.90</u>							
B		100% <i>111</i> <u>-6.30</u>							
C		98.3% <i>293</i> <u>-5.95</u>				1.7% <i>5</i> <u>-5.71</u>			
D		100% <i>18</i> <u>-4.89</u>							

Figure S1. Heat map of the docking results from **Figure 2** in the main text, with the docking scores included. The percentage of low energy structures identified for each docked structure are shown in bold, the number of low energy structures remaining after filtering is shown in italics and underlined is the average interface energy (docking score). The thick black lines separate structures docked with different constraints.

Details of the coordination constraints.

The coordination constraints are the values used to restrain the magnesium/diphosphate complex to the area of the active site observed in crystal structures with a complete active site and all three magnesiums and a diphosphate bound. For BPPS, there are four crystal structures that meet this criteria – 1N20, 1N22, 1N23 & 1N24 on the PDB. The distance, angle and dihedrals from the canonical DDXXD and NSE motifs to their corresponding magnesiums were measured and the average value was taken as the constraint value and their standard deviation was set as a window of deviation from the value for which there would be no penalty.

Table S1. Values for the constraints used to dock the magnesium/diphosphate complex into the relaxed protein structure. The constraints are between the oxygen in the indicated residue and the indicated magnesium in the complex.

D817 constrained to MG1

<i>constraint</i>	<i>value</i>	<i>window</i>
distance	2	0.1
angle	143.3	10
dihedral	197.4	5

D813 constrained to MG2

<i>constraint</i>	<i>value</i>	<i>window</i>
distance	2	0.1
angle	132.9	2
dihedral	51.6	5

D817 constrained to MG2

<i>constraint</i>	<i>value</i>	<i>window</i>
distance	2.3	0.1
angle	125.8	6
dihedral	158.4	5

D958 constrained to MG3

<i>constraint</i>	<i>value</i>	<i>window</i>
distance	2.3	0.2
angle	137.2	11
dihedral	16	25

D813 constrained to MG1

<i>constraint</i>	<i>value</i>	<i>window</i>
distance	1.95	0.1
angle	125.1	5
dihedral	310.4	10

E966 constrained to MG3

<i>constraint</i>	<i>value</i>	<i>window</i>
distance	2.2	0.2
angle	116.6	5
dihedral	241.1	37.4

Figure S2 – How much sampling is required to make a confident prediction?

A large challenge with the stochastic sampling method employed and the large number of degrees of freedom to be sampled, is determination of when sufficient sampling has been conducted. To determine the number of docking runs required to make a confident prediction (i.e., was there convergence?) the first cationic intermediate **A** was docked into the BPPS crystal structure a variable number of times (**Figure S2**). Qualitatively the calculations appear to converge at an nstruc of 100, but such a low number of structures surviving the filtering doesn't lead to a confident prediction, as a random structure in any other orientation would significantly muddy the prediction. Out of an abundance of caution, we ran an nstruc of 2500 for all intermediates in our docking simulation described in the main text and SI.

Number of docking runs	Orientation 1	Ornt. 2	Ornt. 3	Ornt. 4	Ornt. 5	Ornt. 6	Ornt. 7	Ornt. 8
nstruc=100	4	-	-	-	-	-	-	-
nstruc=500	11	-	-	-	-	-	-	-
nstruc=1000	19	-	-	-	-	-	-	-
nstruc=2500	85	1	-	-	-	-	-	-
nstruc=5000	131	-	-	-	-	-	-	-
nstruc=10000	281	1	-	-	-	-	-	-

Figure S2. Convergence tests for BPPS. The number structures in a given orientation that passed filtering is shown for different numbers of docking runs.

Figure S3. Heatmap for the docking results for known stereochemistry of the product into the other crystal structure, 1N23.

In order to ensure our prediction isn't a result of input bias, the docking simulation was performed on an alternate crystal structure (1N23) than that shown in the main text (1N20). The results make the same prediction with only **orientation 1** connecting everything, from **GPP** to all intermediates and the final product. This orientation is still consistent with the labeling experiment by Croteau et al. That both structures come to the same answer gives more confidence in the prediction.





		Orientation							
		1	2	3	4	5	6	7	8
Structure		★	-	-	-	-	-	-	★
		GPP attached to oxygen A				GPP attached to oxygen B			
GPP		91.1% 389				8.9% 38			
A		100% <i>1849</i>	-	-	-	-	-	-	-
B		36.1% <i>364</i>	0.1% <i>1</i>	-	-	63.8% <i>644</i>	-	-	-
C		99.4% <i>876</i>	-	-	-	0.6% <i>5</i>	-	-	-
D		100% <i>47</i>	-	-	-	-	-	-	-

Figure S3. Heat map of the docking results. For GPP, only two discrete orientations were tested: attachment to oxygen A or B. Ion-pair orientations for other structures are depicted in **Figure 2** in the main text. The darker the color filling each cell of the table, the more low energy solutions were found for that orientation. The percentage of low energy structures identified for each docked structure are shown in bold, the number of low energy structures remaining after filtering is shown in italics. The thick black lines separate structures docked with different constraints.

Figure S4. Violin plot to identify pathway of least movement for the docking results for 1N23 and the observed stereochemistry.

The docking results from 1N23 were subjected to the same RMSD analysis that the results in the main text were. The results are plotted below in **Figure S4**.

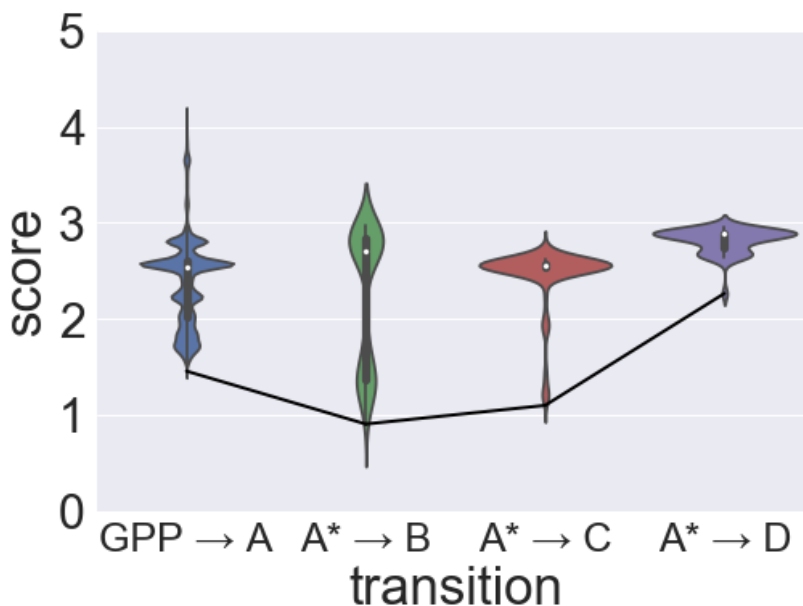


Figure S4. Violin plot of the RMSD for each transition in the other BPPS crystal structure, 1N23. The mean value is indicated as a white dot and the interquartile range as a thick black box around the mean, the 95% confidence interval are shown as a black line. The population at any given RMSD score is mirrored on both sides. A line connecting the lowest RMSD structure identified in the search has been drawn in black.

Figure S5. Overlay of the pathway identified by the violin plots

The poses identified in the RMSD search in **Figure S4** are shown here. Again, the ligand converges to a single region in space (**E**) and is predicted with the same orientation as the crystal structure and with the correct stereochemistry (**F**)

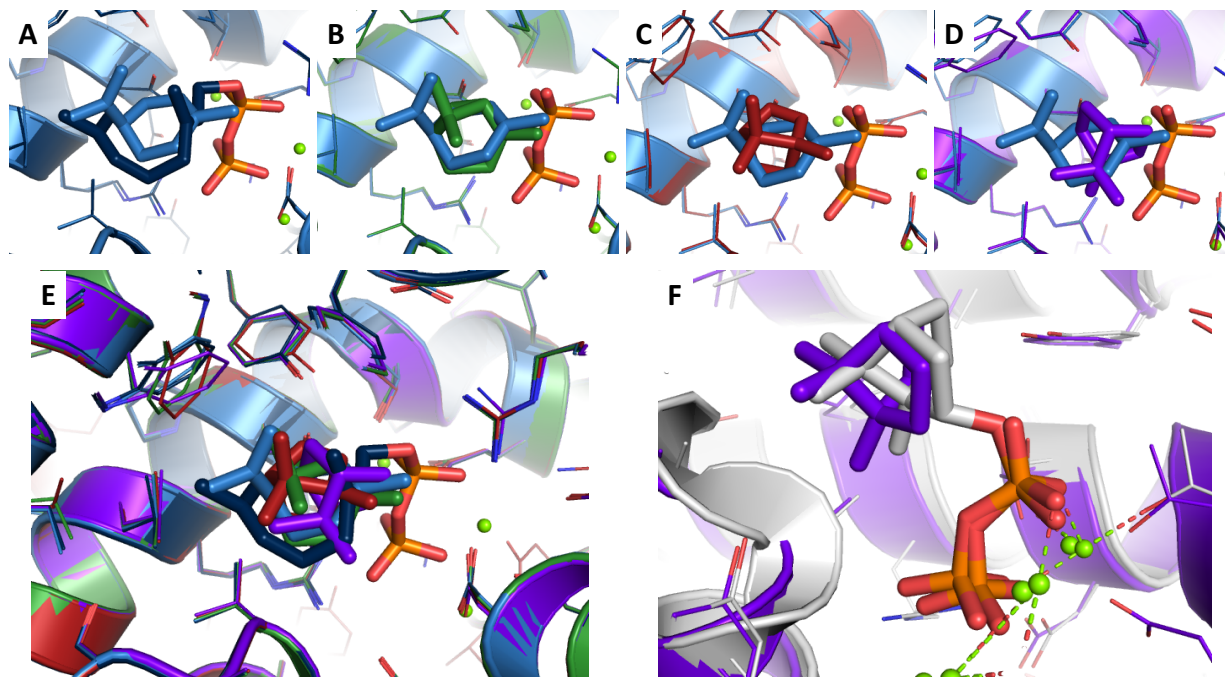


FIGURE S5. **(A)** Overlay of GPP (navy) and intermediate **A** (blue), **(B)** Overlay of intermediate **A** and intermediate **B** (green), **(C)** overlay of intermediate **A** and intermediate **C** (red), **(D)** overlay of the intermediate **A** and the product, **D** in purple. **(E)** Overlay of all identified structures together. **(F)** The orientation of the product shown in purple overlaid with the crystal structure with the product bound in white (PDB 1N24).

Figure S6. Heatmap for the docking results for enantiomer stereochemistry of the product into the other crystal structure, 1N23.

To ensure that the enantiomer docking also wasn't a result of input bias the same docking simulation was performed into the crystal structure 1N23. Like the results described in the main text, no orientation connects all the structures modeled. Also, like the results described in the main text, the number of structures in *ent-A* is significantly reduced compared to that for **A** (see **Figure S3** above).

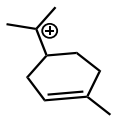
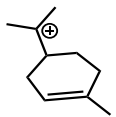
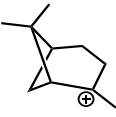
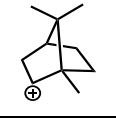
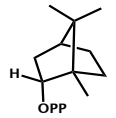
		Orientation							
		1	2	3	4	5	6	7	8
Enantiomer Structure		★	-	-	-	-	-	-	★
		GPP attached to oxygen A				GPP attached to oxygen B			
GPP		88% 381				12% 52			
A		72.4% <i>21</i>	-	-	27.6% <i>8</i>	-	-	-	-
B		88.2% <i>829</i>	-	-	7.9% <i>74</i>	0.1% <i>1</i>	-	-	3.8% <i>36</i>
C		100% <i>98</i>	-	-	-	-	-	-	-
D		-	-	-	-	-	-	-	-

FIGURE S6. Heatmap of the results for docking the enantiomer of the carbon skeleton into the active site of BPPS (PDB 1N23). Each orientation is described in **Figure 2** in the main text. The orientations highlighted with a star are consistent with the labeling experiment conducted by Croteau et al. The darker the color the more low energy solutions are in that orientation. The number of low energy structures after filtering are in bold and the average interface energy (docking score) for that orientation are in italics. The number in red for intermediate A is to indicate a significantly lower number of solutions identified for the enantiomer compared to the stereochemistry observed for the product.

Figure S7 & S8. Details on alternate filtering on the results for the *ent*-D docking.

When no structures of **ent-D** passed the standard filtering criteria used for the other structures, an investigation was made into what would be required in order for a viable binding orientation to be predicted. If the filtering criterion for the constraints was relaxed to 2 (from 1) for the simulations in 1N20, and after the regular filtering method is applied, only eight structures survive the process and all of those low energy structures are in **orientation 5** (Figure S7). In order to get structures after filtering for 1N23, the constraint criteria was relaxed to 2.5 (again from 1) and the results are nearly identical to the modeling in 1N20 – very few structures pass, despite the more permissive filter and the structures that do are in **orientation 5** (Figure S8). This illustrates that the lack of a pathway that connects all intermediates is not filtering dependent.

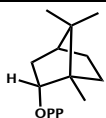
		Orientation							
		1	2	3	4	5	6	7	8
Enantiomer Structure		★	-	-	-	-	-	-	★
		GPP attached to oxygen A				GPP attached to oxygen B			
D		-	-	-	-	8 100%	-	-	-

Figure S7. 1N20 *ent*-D docking results with more relaxed filtering.

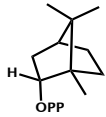
		Orientation							
		1	2	3	4	5	6	7	8
Enantiomer Structure		★	-	-	-	-	-	-	★
		GPP attached to oxygen A				GPP attached to oxygen B			
D		-	-	-	-	14 100%	-	-	-

Figure S8. 1N23 *ent*-D docking results with more relaxed filtering.

Discussion on enantiomer formation from alternate substrate

(+)-Bornyl diphosphate synthase (BPPS) from sage makes enantiomerically pure (+)-BPP from geranyl diphosphate (GPP) through what is thought to be (3*R*)-linalyl diphosphate (LPP). If the enzyme is fed the enantiomer (3*S*)-LPP, then the same enzyme will make the enantiomeric product (-)-BPP.¹ Currently, the *TerDockin* method is unlikely to provide good insight about this very interesting experimental result. This is likely a sampling issue, with only 18 structures for the known (+)-BPP stereochemistry passing the filtering and no (-)-BPP passing the filtering at the same level of sampling the number of docking runs would likely have to increase by a minimum of an order of magnitude over the currently being produced 20,000 structures per intermediate, in order to make a meaningful prediction. Unfortunately, we feel that this makes this analysis computationally intractable.

Table S2. Energies of the Iodo-replacement conformers.

Conformer No.	File name	Free Energy (a.u.)	Relative Energy(kcal/mol)
1	1.mol2	-402.375768	0
2	7.mol2	-402.375257	0.32065761
3	2.mol2	-402.374425	0.84274593
4	8.mol2	-402.374286	0.92996982
5	10.mol2	-402.374252	0.95130516
6	5.mol2	-402.374067	1.06739451
7	6.mol2	-402.373744	1.27008024
8	9.mol2	-402.373707	1.29329811
9	4.mol2	-402.373468	1.443273
10	16.mol2	-402.37335	1.51731918
11	12.mol2	-402.373196	1.61395572
12	14.mol2	-402.372873	1.81664145
13	13.mol2	-402.37287	1.81852398
14	15.mol2	-402.372785	1.87186233
15	17.mol2	-402.372497	2.05258521
16	11.mol2	-402.371812	2.48242956
17	18.mol2	-402.371376	2.75602392
18	3.mol2	-402.3713	2.80371468

The structures identified in the conformation search have also been provided as mol2 files.

Figure S9. Heatmap of docking into 1N20 crystal structure with new score function in Rosetta, ref2015

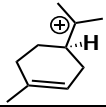
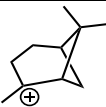
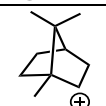
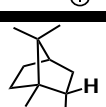
Structure		Orientation							
		1	2	3	4	5	6	7	8
		☆	-	-	-	-	-	-	☆
		GPP attached to oxygen A				GPP attached to oxygen B			
GPP		100% 214				-			
A		100% 694	-	-	-	-	-	-	-
B		97.4% 189	-	-	-	2.6% 5	-	-	-
C		100% 471	-	-	-	-	-	-	-
D		100% 6	-	-	-	-	-	-	-

FIGURE S9. Heatmap of the results for docking the carbon skeleton into the active site of BPPS (PDB 1N20), using the newer Rosetta score function. Each orientation is described in **Figure 2** in the main text. The orientations highlighted with a star are consistent with the labeling experiment conducted by Croteau et al. The darker the color the more low energy solutions are in that orientation. The number of low energy structures after filtering are in bold and the average interface energy (docking score) for that orientation are in italics.

Recently, a new score function for Rosetta was published.² To ensure the results obtained in the main text are not score function dependent, the docking for structures was redone with the new score function and the results are show in **Figure S9**. The new score function, called ref2015, does not change the prediction made with Rosetta when using the older score function, called talaris2013.³⁻⁵ This provides evidence that the prediction made by the docking simulations are not dependent on the score function.

References

- (1) Croteau, R.; Satterwhite, D. M.; Cane, D. E.; Chang, C. C. *J. Biol. Chem.* **1986**, *261*, 13438-13445.
- (2) Alford, R. F.; Leaver-Fay, A.; Jeliazkov, J. R.; O'Meara, M. J.; DiMaio, F. P.; Park, H.; Shapovalov, M. V.; Renfrew, P. D.; Mulligan, V. K.; Kappel, K.; Labonte, J. W.; Pacella, M. S.; Bonneau, R.; Bradley, P.; Dunbrack, R. L.; Das, R.; Baker, D.; Kuhlman, B.; Kortemme, T.; Gray, J. *J. J. Chem. Theory and Comput.* **2017**, *13*, 3031-3048.
- (3) Leaver-Fay, A.; O'Meara, M. J.; Tyka, M.; Jacak, R.; Song, Y.; Kellogg, E. H.; Thompson, J.; Davis, I. W.; Pache, R. A.; Lyskov, S.; Gray, J. J.; Kortemme, T.; Richardson, J. S.; Havranek, J. J.; Snoeyink, J.; Baker, D.; Kuhlman, B. In *Method Enzymol.*; Keating, A. E., Ed.; Academic Press: 2013; Vol. 523, p 109-143.
- (4) Shapovalov, Maxim V.; Dunbrack, Roland L. *Structure* **2011**, *19*, 844-858.
- (5) Song, Y.; Tyka, M.; Leaver-Fay, A.; Thompson, J.; Baker, D. *Proteins: Struc., Funct., Bioinf.* **2011**, *79*, 1898-1909.